5-2016

Machine Learning and Neutron Sensing in Mobile Nuclear Threat Detection

John Edward Sparger
University of Tennessee - Knoxville, jsparger@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by John Edward Sparger entitled "Machine Learning and Neutron Sensing in Mobile Nuclear Threat Detection." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Nuclear Engineering.

Jason P. Hayward, Major Professor

We have read this dissertation and recommend its acceptance:

Lawrence Heilbronn, Hairong Qi, Howard Hall

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Machine Learning and Neutron Sensing in Mobile Nuclear Threat Detection

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John Edward Sparger

May 2016

Acknowledgements

This work has been supported by the U.S. Department of Homeland Security under competitively awarded Interagency Agreements HSHQDC-10-X-00662 and HSHQDC-13-X-00172. This support does not constitute an express or implied endorsement on the part of the government.

Abstract

A proof of concept (PoC) neutron/gamma-ray mobile threat detection system was constructed at Oak Ridge National Laboratory. This device, the Dual Detection Localization and Identification (DDLI) system, was designed to detect threat sources at standoff distance using neutron and gamma ray coded aperture imaging. A major research goal of the project was to understand the benefit of neutron sensing in the mobile threat search scenario. To this end, a series of mobile measurements were conducted with the completed DDLI PoC. These measurements indicated that high detection rates would be possible using neutron counting alone in a fully instrumented system. For a 280,000 neutrons per second Cf-252 source placed 15.9 meters away, a 4σ detection rate of 99.3% was expected at 5 m/s. These results support the conclusion that neutron sensing enhances the detection capabilities of systems like the DDLI when compared to gamma-only platforms.

Advanced algorithms were also investigated to fuse neutron and gamma coded aperture images and suppress background. In a simulated 1-D coded aperture imaging study, machine learning algorithms using both neutron and gamma ray data outperformed gamma-only threshold methods for alarming on weapons grade plutonium. In a separate study, a Random Forest classifier was trained on a source injection dataset from the Large Area Imager, a mobile gamma ray coded aperture system. Geant4 simulations of weapons-grade plutonium (WGPu) were combined with background data measured by the Large Area Imager to create nearly 4000 coded aperture images. At 30 meter standoff and 10 m/s, the Random Forest classifier

was able to detect WGPu with error rates as low as 0.65% without spectroscopic information. A background subtracting filter further reduced this error rate to 0.2%. Finally, a background subtraction method based on principal component analysis was shown to improve detection by over 150% in a figure of merit.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the wake of September 11th, 2001, the United States of America began a large scale campaign to enhance its national security. As part of this effort, an arsenal of nuclear security technologies and polices were developed, each designed to close a gap in the nation's defense. In this document, we will be concerned with one gap scenario in particular: "If a nuclear weapon or dirty bomb was smuggled into a U.S. city, could it be detected?"

In the years since 9/11, this question has received significant attention. A number of mobile search platforms have been developed within the last decade which attempt to provide this capability. These vehicle mounted systems use advanced radiation detectors and imaging technology to locate sources at a distance. And, since they conduct measurements on the move, large areas can be scanned in a short amount of time.

Mobile search is not, however, a solved problem. The variability of natural background radiation, shielding, and weak source signatures serve to make the detection problem a difficult one. While imaging and spectroscopy can mitigate some background variation effects, shielding and weak sources are fundamental limitations in the problem space. No amount of cleverness will detect a completely shielded source.

In this document, the development of the Dual-Detection-Localization-Identification (DDLI) system is described. The DDLI approached mobile search in two unique ways. The first was systematic. Previous mobile imaging platforms had focused on gamma ray sensing. However, many special nuclear materials have weak gamma signatures or can be shielded by just a few centimeters of heavy metal. By incorporating neutron detectors, the DDLI sought to detect neutron emitting sources whose gamma signature was too weak. The second unique contribution came in the form of algorithm development. While a mobile detector platform generates a significant amount of information, leveraging this information effectively can be a problem. In the DDLI studies, machine learning techniques were used to combine data, subtract background, and improve the detection bottom line.

In Chapter 2, the research topics of this dissertation and the DDLI project at large will be outlined. Concepts important to the entire document will then be discussed in Chapter 3. Chapter 4 details machine learning studies aimed at improving the alarm algorithms which ultimately detect sources. An introduction to the topic of machine learning and a review of related works in literature are included. Chapter 5 covers the build, configuration, and testing of a prototype DDLI system. Information on other mobile imagers and the early DDLI project work are provided as context. Finally, conclusions and ideas for future work are laid out in Chapter 6.

Chapter 2

Problem Statement

2.1 Project Topics

The DDLI project was proposed in response to the Domestic Nuclear Detection Office (DNDO) CPF10-DNDO-01 TA 2.4, "Technologies to Support Next Generation Dual Neutron/Gamma Imaging." This far reaching topic encompassed many research areas:

1. Simultaneous gamma ray and fast neutron imaging and spectroscopy
2. New materials sensitive to both fast neutrons and gamma rays.
3. Signal processing for neutron and gamma ray separation
4. Advanced electronics for sensing the position and time correlation of interactions
5. Mobile or transportable designs
6. Advanced algorithms to improve detection performance
7. Faster running algorithms for real-time imaging
8. Methods to discriminate signal from background clutter
9. Ability to support active detection
10. Hybrid designs incorporating multiple detector types

Nearly all of these topics were addressed during the course of the project. Topic (1) was addressed through the use of fast organic scintillators. Item (2) was investigated by performing simulation studies of new materials like CLYC and Elpasolites. All system designs were mobile hybrid imagers, as per (5) and (10). An effort to improve detection performance through machine learning and image processing is described in detail in Chapter 4 of this document. This study was conducted to target (6) and (8). Finally, item (9) was fulfilled by examining the utility of including of thermal neutron detection technology into the system.

Along with the broad research areas of TA 2.4 came specific project goals:

1. One percent energy resolution for gamma rays at 662 keV
2. Position sensitivity with 1-degree angular resolution
3. Detection of neutron or gamma emitting sources of concern at 100 meters in operationally relevant time spans
4. A proof of concept (PoC) prototype demonstration

Knowledge of these project requirements will illuminate some of the design choices presented in this document. As is to be expected, the project evolved as each area of interest was explored in more detail.

2.2 Problem Statement

The work of this dissertation is concerned with two of the broad areas in the DDLI project. The first is the development of the DDLI PoC prototype. The second is the development of advanced algorithms.

2.2.1 Proof of Concept Build

As mentioned in Section 2.1, a PoC prototype demonstration was an expected deliverable for the DDLI project. At the time that the work of this dissertation

began, the DDLI system had been thoroughly characterized in simulation space. What remained was to demonstrate real-world performance with a physical system. The novelty of the DDLI design was neutron detection. PoC demonstrations would therefore be focused on characterizing the benefits of neutron sensing in mobile search.

The task of the author would be to build, configure, and test a mobile code aperture imager. Then, a series of measurements would need to be designed to test whether neutron sensing could enhance detection capabilities in a practical way. After the execution of this measurement campaign, the PoC data would need to be collated and processed to make final performance reports.

2.2.2 Algorithm Development

Algorithms are an extremely important part of a threat detection system. The physical construction of an instrument fundamentally determines its capabilities. However, it is the algorithms which determine how well these capabilities may be leveraged. The most advanced platform is useless without an effective way to process its data.

In the call for proposals, advanced algorithm development and methods to separate signal from background were specifically requested. One response to this call would be to investigate data fusion. To make the most of the DDLI's hybrid design (see Section 5.3.1), signals from different detector types and imaging modalities would need to be integrated into a single alarm. The task, then, was to find methods which generated alarms from disparate data.

2.3 Original Contributions

The original contributions of this work are briefly summarized as follows:

1. The build, configuration, and testing of a mobile, neutron/gamma-ray coded aperture imaging system (Chapter 5)

2. Demonstration of added benefit from neutron sensing in mobile threat detection (Section 5.7.2)

3. Fusion of 1-D gamma ray and neutron coded aperture image data using machine learning (Section 4.5)

4. Demonstration of performance benefits from machine learning alarm algorithms in gamma ray coded aperture imaging (Section 4.6.5)

5. Suppression of background in coded aperture images through a spatial frequency filter (Section 4.6.6)

6. Subtraction of background from coded aperture images through a generative PCA method 4.6.7)

Chapter 3

Important Concepts

3.1 Mobile Counting

If we wish to determine whether a source is present near a stationary detector, we may conduct a counting experiment. In this experiment, we will simply count the number of interactions we receive in our detector over some finite period of time. If the number of counts exceeds some predetermined threshold, we will issue an alarm for detection.

Clearly, the important parameter in this alarm scheme is the detection threshold. How should it be set? If we can make a measurement without the source present, we can establish a mean background count rate for our location. Since nuclear counting is a Poisson process, the mean count rate fully defines the background counting distribution. Knowing the mean count rate, we can set a threshold to some value above the mean, say 5σ, which will give our alarm a sufficiently low false positive rate. Now, whenever a source is introduced, if the combined source plus background count rate exceeds our threshold, we can alarm with high confidence, as it is statistically unlikely that background alone could exceed the threshold.

The simple counting technique presented above can often be very effective. Importantly, though, it makes a few key assumptions about the nature of the

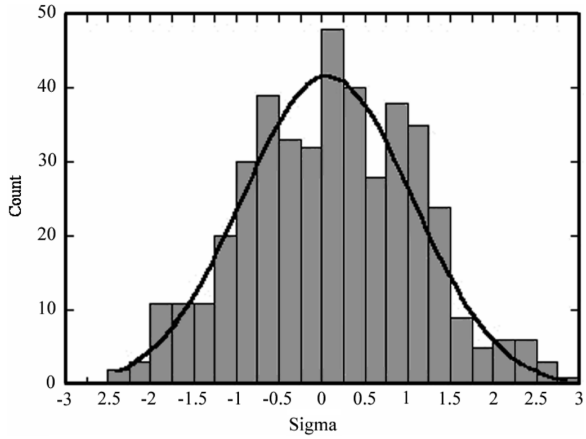


Figure 3.1: Expected distribution shape for constant background. [1] © 2006 IEEE.

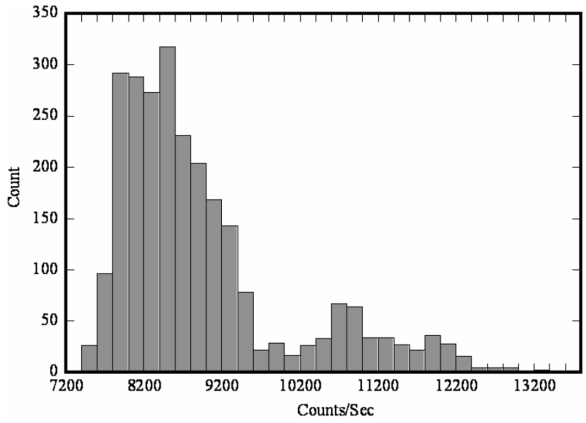


Figure 3.2: Background counting distribution measured during mobile runs of the Large Area Imager around Livermore, California. [1] © 2006 IEEE.

background signal. For this scheme to work, the mean background count rate must be constant. Unfortunately this is not even approximately the case in a mobile gamma ray measurement.

Figure 3.1 shows the expected shape of the counting distribution when the true rate is constant. This is in stark contrast to the counting distribution shown in Figure 3.2, which was measured by the Large Area Imager (LAI) during a mobile, constant-speed background measurement around Livermore, California. [1]

The large, non-Poisson background fluctuations in Figure 3.2 can be attributed to changes in the environment as the imager drove around Livermore. Sources of gamma background radiation, such as buildings, roads, and the Earth's crust change

in composition, proximity, and density as the imager moves. The imager can drive over and under bridges; it can drive through urban and rural areas; other vehicles can shield the imager as they pass by. The combination of these factors leads to a highly variable mean count rate which is heavily correlated with the position and orientation of the imager.

If we attempted to apply our Poisson threshold based on the mean count rate of the mobile data taken by the LAI, we would underestimate the variance of the count rate and see an unexpectedly large number of false alarms. If we increased the threshold until we achieved an acceptable false alarm rate, the count rate needed to trigger an alarm would be very high. With such a threshold, we would likely not detect sources of interest in low background areas.

There are some methodologies to deal with this problem, such as using sliding windows to estimate the instantaneous mean background count rate [12]. These windows must of course be set to a range which reflects the temporal frequency of background variation, which may be difficult to quantify and may need to be adjusted dynamically. There are also statistical considerations when considering the window length, as fewer counts will lead to a greater uncertainty. Naturally, these schemes may become as complicated as necessary and can yield better results. However, in a basic sense, this type of analysis is an attempt to add spatial information into the alarm algorithm. In Section 3.2, we will discuss a more direct way to incorporate this information.

3.2 Coded Aperture Imaging

Coded aperture imaging is an extension of the pinhole camera. The first coded aperture systems replaced the single pinhole with many pinholes in some pattern, often a random one. The goal of this approach was to increase the amount of signal reaching the detector without increasing the aperture of the pinhole, which would reduce the angular resolution. The measured image is thus a superposition of the

various pinhole images of the source. In order to reconstruct the object image, a decoding step is performed which is typically implemented as a correlation. The following equations detail the measurement and reconstruction process:

$$P = (O * A) + N \tag{3.1}$$

$$\widehat{O} = P * G \tag{3.2}$$

where P is the measured image, O is the object, A is the coded aperture array, N is noise, $\widehat{O}$ is the reconstructed object, G is some reconstruction array, and $*$ is the correlation operator.

A common choice for the mask pattern is a uniformly redundant array (URA) because it has a system point spread function which is effectively a delta function. URA's have several advantages over random arrays. One advantage is that the signal to noise ratio can improve even when imaging non-point sources. Another advantage is the ability to tile the pattern so that the detector need not be large enough to detect the entire shadow to perform reconstruction. [13]

A URA image may be reconstructed by performing a "balanced correlation" [13], in which the G of equation (3.2) takes the form

$$G(i) = \begin{cases} 1, & \text{if } A(i) = 1 \\ -1, & \text{if } A(i) = 0. \end{cases} \tag{3.3}$$

Chapter 4

Machine Learning

4.1 Introduction

While the sensing components of a nuclear threat detection system are of obvious, great importance, they are of no use without some method of analyzing their outputs to recognize the threat. In this sense, detection systems, no matter how complicated, are eventually reduced to a single binary signal indicating the presence or absence of a source. This signal, the alarm, is the final decision in a threat source detection scenario.

A useful alarm signal must be correlated to the actual presence of a threat source. A perfect signal would be perfectly correlated: always alarming when a threat was present and never alarming when no threat was present. Of course, the limitations of our sensors will mean that we must live with some number of false alarms (false positives) and real threats which elude detection (false negatives). The goal in designing the alarm signal is to achieve a true positive rate as close as possible to 100% while keeping the false positive rate at a level that is manageable in practical operation.

Mobile nuclear threat detection systems of the type under consideration in this work are complicated (see section 5.2). They may consist of hundreds of detectors

of varying types arranged such that they work individually or in concert to perform various imaging tasks. They may detect gamma-rays, neutrons, or both, the final case requiring the separation of neutron and gamma induced signals. The detector signals may also contain spectroscopic information describing the energy of detected particles. Other sensors may indicate the speed at which the system is moving or its GPS coordinates in the world. Still more sensors may give the current temperature, humidity, and weather.

Considering that each sensor or sensing modality may contribute valuably to the detection of a source, we are inclined to include as many of the signals at our disposal into the alarm signal as we can. We must then come up with a set of rules to determine what combinations of signals will result in an alarm and which combinations we will deem to be normal. Although simply stated, this is in fact a daunting task. Not only do we need to combine signals which may be related in complicated ways, we must also take into account the wide variety of sources we are interested in detecting and how their signatures are affected by distance, shielding and dwell time. Perhaps equally as challenging, we must incorporate the effects of fluctuating background which varies as a function of our location and other environmental variables.

While we may be able through great effort and intuition to construct a set of rules addressing many of the conditions listed above, how can we be convinced that our solution is optimal – or at least that it is not severely suboptimal? With such a large set of parameters to tune, we may miss some important relationships between signals. In some cases, our intuition may be flawed. In other cases we may have no intuition as to how to meaningfully relate some subsets of variables, even though a meaningful relationship may exist.

An alternative in this case is to attempt to learn detection rules from data. If we possess enough examples, can we recognize informative patterns within our high dimensional signal data? Can we recognize certain combinations of signals which are indicative of threat sources and distinguish them from combinations which are indicative of background? Fortunately, a wide variety of machine learning algorithms

have been developed which can perform this task. These machine learning classifiers can learn to associate labels, such as 'source' or 'background', with regions of a high dimensional space by examining training datasets where the correct answer is known. This is known as supervised classification. Still other algorithms can help us transform the individual signals we are working with such that they are condensed to their most discriminatory components or features. Even when we have hand selected the representations of our signals (as opposed to learning them), we can use optimization techniques and estimators of variable importance to prune our feature set. In this way we may reduce the complexity of our problem by eliminating signals which are not useful. See Section 4.4.1 and Section 4.4.3 for a detailed discussion of these techniques.

4.2 Overview

This portion of the document is concerned with the application of machine learning techniques to nuclear threat detection. Section 4.3 presents examples of related work from literature. The theory of machine learning algorithms relevant to this work is then discussed in 4.4. Finally, a series of studies are presented. The first study in Section 4.5 concerns machine learning for data fusion in a simulated mobile imaging dataset. The second study in 4.6 details machine learning and background subtraction in a source injection study with measured background. A description of software developments for the project may be found in Appendix A.

4.3 Literature Review

Advanced machine learning has only recently seen an increase in application with respect to nuclear threat detection. Many of the advanced techniques employed in other fields have yet to spill over into this problem area, especially those which have developed in the computer vision community. In this section, we highlight some

examples from the literature where machine learning techniques have been applied to threat detection scenarios. In some cases, real world threat detection systems have been developed based on these ideas. In other cases, machine learning has been applied to specific aspects of the threat detection problem such as gamma spectroscopy.

4.3.1 Bayesian Aggregation

Carnegie Mellon's Robotics Institute has produced a Bayesian Aggregation (BA) framework for detecting, localizing, and identifying threat sources using mobile detector measurements under varying background [14]. This framework relies on generating probability distributions parameterized by an exposure statistic and a signal-to-noise ratio (SNR) score. These 2D probability density distributions represent the likelihood that a given exposure-SNR pair will be observed given a set hypothesis. By referencing a probability distribution for both the case where a threat source is present as well as the case where a threat source is not present, BA can evaluate the likelihood that a given measurement was generated from each model. With this likelihood in hand, a posterior probability can be computed by multiplying it with some prior. In the end, the model with the highest posterior probability is chosen as the true model.

In the Bayesian Aggregation framework, each measurement may affect the detection decision at many different spatial locations. To aggregate the measurements, BA assumes complete independence. Thus, the posterior for an aggregated measurement is calculated according to

$$P(H \mid s(D)) \propto P(H) \prod_{D_j \in D} P(s(D_j) \mid H). \tag{4.1}$$

In this equation, $P(H \mid s(D))$ is the probability of hypothesis H given the SNR scores $s(\cdot)$ for the series of measurements D. The quantity $P(H)$ is the prior probability

assigned to the hypothesis, and $P(s(D_j) \mid H)$ represents the likelihood of observing the score observed from the jth measurement. Because measurements are assumed to be independent, the likelihoods for each measurement can be multiplied together to obtain the overall likelihood.

Two SNR scores are discussed in [14]. These scores are an "anomaly score" based on PCA residuals and a "matched filter" based on a least squares template matching. Both scores are calculated using measured spectra which have been redistributed into 128 energy bins. Each SNR score is computed by first estimating the individual contributions from source and background to a given measurement. Once the estimates have been obtained, the SNR is calculated according to

$$SNR = \frac{S}{\sqrt{B}} \tag{4.2}$$

To compute the anomaly score, Principal Component Analysis (PCA, see Section 4.4.3.1) is performed using measured background spectra. An unknown measurement is then projected onto the first five principal components. The sum of the counts from this projection serves as the background contribution estimate, B, while the L_2 norm of reconstruction error serves as the source contribution estimate S.

For the matched filter score, spectrum templates are reduced to a subset of their 128 bins. These bins are selected such that they provide the best SNR, i.e. they are the bins which most differentiate a given source from the mean background spectrum. Using a regression estimator trained to predict the background content in the selected bins based on the counts in the unselected bins, the source and background estimates for the selected bins can be obtained.

The exposure statistic is computed from vehicle velocity, hypothesized source intensity, and relative locations of the source and vehicle. This statistic is

representative of how strong a source should appear and is used as a multiplier for injected source counts. This multiplier could also be used to account for known occlusions.

Probability density functions are computed with SNR scores and exposure statistics for a training data set. This set of training data pairs is then used in conjunction with non-parametric density estimators to estimate the underlying distributions.

In source injection tests, BA was able to both localize sources to within 40m and give some estimation of the source intensity. In a test with three different source types, BA was also able to infer the correct source type with high accuracy.

4.3.2 Portal monitors and Random Forest

In 2012, Simon Labov presented two posters [15][16] detailing the use of Random Forest (see Section 4.4.2) for improving the performance of radiation portal monitors. A large feature set was extracted from thousands of cargo measurements, with sources simulated via injection. One subset of features was statistical moments taken from the count rate, energy, and position distributions generated by a scan (the position and count rate distributions exists due to the fact that cargo is driven through the portal monitor). A second subset was taken from PCA features of the portal monitor scan trace while a third was extracted by removing high and low frequency components from the scan trace. A final subset was taken from goodness-of-fit metrics and the fit parameters for four scan trace source models. These models represented non-radioactive, compact, distributed, and compact plus distributed sources.

Using this setup, Labov showed that a random forest classifier could outperform an optimized spectral analysis, a common handheld detector model, and a scan based on the number of standard deviations above background. He also showed that random forest could continue to outperform the standard-deviations-above-background approach in the case where no energy spectrum information was available.

Finally, he demonstrated that different classes of materials, such as naturally occurring radioactive material (NORM) and threat sources, were distinguishable and showed good separation in the feature space.

Interestingly, a portal monitor study had already been used to test the performance of a random forest variant two years earlier in a data mining paper by Lemmond [17]. The work concerned random forest and a variant called discriminant random forest on a dataset of 128 bin energy spectra. These spectra, taken from real vehicles passing through a portal monitor, were subjected to source injection with uranium and plutonium. Though the details of the dataset are somewhat unclear, the author used a very large training dataset (>10k samples) to demonstrate high detection accuracy with both methods. In this study, discriminant random forests significantly improved upon the standard random forest results.

4.3.3 Spectral Background Subtraction

The RadMAP platform (see Section 5.2.1.4) has been used to create an extensive dataset of background measurements from around the San Francisco Bay Area in California. This dataset contains measurements from HPGe detectors and also NaI detectors behind a lead 2D coded aperture mask. In a paper by Aucott, RadMAP background data was used to perform a NaI source injection study at 100m standoff to determine the relative benefits of imaging vs. spectroscopy [6]

Aucott shows that variations in the gamma ray background caused the ^{137}Cs photopeak counts to exhibit a variance over 100 times greater than Poisson fluctuations alone. This high variance leads to almost nonexistent detection performance. It is suggested that estimating and subtracting the background counts within the photopeak window would decrease this variance substantially. Two methods for subtracting this background are suggested. One is a simple linear interpolation based on the values of neighboring regions in the spectrum. The second is a three parameter linear estimate based on neighboring regions and a peak excess

value. The parameters of this model are learned from a separate set of background training data using least squares regression. In his final results, Aucott demonstrates vastly improved detection performance in the spectroscopic case using the background subtraction methods, with the trained model outperforming the interpolation.

4.3.4 Dimensionality reduction for detection and spectral ID

An intuitive method for computing an anomaly detection metric is to employ PCA (Section 4.4.3.1) on energy spectra. Intuitively, this is because the principal components capture the orthogonal directions which explain the most variation within a given dataset. These directions, especially the first, represent the directions of highest correlation, which is an effective representation of the relationship between energy bins. If PCA is performed on a background only dataset, one would expect the first few principal components to do a good job representing other background measurements. A source spectrum, however, would be expected to reconstruct poorly from these components due to the existence of extra trends not captured during their generation (see Section 4.3.1).

In [18], Boardman proposes a PCA based anomaly detection algorithm for radiation portal monitors. This algorithm performs PCA on a training set of background spectra using the correlation matrix (as opposed to the covariance). Then, the first few principal components are used to project a series of reference background spectra. When an unknown point is to be characterized, it is projected into the principal component space and its Mahalanobis distance from the reference spectra is calculated. An alarm is raised when the Mahalanobis distance exceeds a set threshold, which is the distance range within which NORM sources typically fall.

Using this PCA algorithm, Boardman shows that scanning speeds to reach ANSI N42.35 standards for a test track portal monitor could be improved by a factor of 2-4 times for over results obtained with a commercially available peak search algorithm.

He also shows that the PCA algorithm is sensitive to several threat sources which were not detectable using the commercial algorithm, including highly enriched uranium (HEU) and weapons grade plutonium (WGPu).

In a second paper, Boardman employs Fisher's Linear Discriminant (FLD) for spectral ID [19]. Using the same dataset as in the PCA paper, 13 one-vs-all FLD directions are computed using training data containing 13 different sources including background. Similar to the PCA algorithm, the Mahalanobis distance between a test point and the cluster generated by a set of reference spectra is used as a metric. In this case, however, the reference spectra consist of examples from all the other classes. This test is performed for each FLD direction. When the Mahalanobis distance exceeds a certain threshold, a detection for the FLD direction's associated nuclide is registered.

Boardman compares his results again agains a commercial peak finding algorithm. This time the algorithm is named to be IDENTIFY from GBS Elektronik GmbH. The FLD algorithm again displays superior performance of the commercial algorithm. Additionally, it shows consistently increasing performance with increased measurement time, which is not the case for the peak-finding algorithm.

4.4 Machine Learning

4.4.1 Classification algorithms

A classifier is a device which associates a vector of inputs with an output class label. For our purposes, we will concentrate on the binary case where the label set contains only two values: source or background. However, we may wish at some point to make a distinction between different types of sources, in which case our label set can contain an identifier for each category of source. The inputs to the classifier are numbers extracted from the system's data stream. In this discussion, we will refer to these numbers as "features".

This section will provide a general overview of some of the most commonly used machine learning classifiers. The core concepts relating to most of these classification algorithms are simple to understand. Nevertheless, these classifiers are being used to achieve state of the art performance on a variety of real world problems.

A discussion of these algorithms is important so that the results presented in Sections 4.5 and 4.6 can be thoroughly understood. Many of these algorithms were also employed in the related works reviewed in Section 4.3.

4.4.1.1 Bayesian methods

Bayes' rule can be stated for the classification problem as:

$$P(\omega_j|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_j)P(\omega_j)}{p(\boldsymbol{x})} \tag{4.3}$$

where $\boldsymbol{x}$ is the datapoint (a vector of feature values) and ω_j is the jth class. The quantity $P(\omega_j|\boldsymbol{x})$ is referred to as the *a posteriori* probability, or simply posterior. The likelihood is defined as$p(\boldsymbol{x}|\omega_j)$. Finally, $P(\omega_j)$ is termed the *a priori* probability, or prior [20].

Bayes' rule represents an exact, closed form representation of all the knowledge we possess about a given classification problem. With enough information, the ideal classifier may be constructed around this simple expression. Maximum likelihood (ML), where we seek the class that maximizes the likelihood function, and maximum *a posteriori* (MAP), where we seek the class which maximizes the posterior, are common classification schemes based on Bayes' rule. In the case where the priors and likelihoods are known exactly, the MAP classifier has the has the minimum possible error rate! [20]

Unfortunately, it is often difficult in practice to obtain the quantities required for a Bayes classifier. The likelihood can be estimated from a dataset, but an accurate estimation is not trivial and sometimes simply not possible. This is especially true when data is limited or the dimensionality is large. Knowledge of the underlying

distribution's form can help in these cases. Often, simplifying assumptions can be made to fit the problem into the context of Bayes' rule, such as assuming normally distributed variables.

4.4.1.2 K-nearest neighbor

K-Nearest Neighbor (KNN) is a nonparametric classification method [21] for with optimal properties (in the infinite limit) [20]. Here, nonparametric refers to the fact that a parameterized description of the class distributions is not required. In the case of k-Nearest Neighbor, we can proceed directly from the data to a decision without explicitly taking the global form of the distributions into account. In many cases, non-parametric methods provide the only viable solution for performing classification with a dataset. Sometimes we do not know the underlying distribution of the data. This may be because we do not have enough data to determine the parameters, or because the actual form of the distribution is too complicated to identify.

K-Nearest Neighbor is also an example of instance based learning [22]. This means that the generation of the decision boundary is delayed until the actual time of classification. This can be advantageous in the case where we may want to add or subtract training data on the fly to refine the classification. Because KNN considers the training data directly at the time of classification, no internal model needs to be altered to accommodate these changes. The disadvantage of this approach is that the computational burden is shifted to classification time, meaning that KNN can be quite a bit slower than a classifier which precomputes a decision boundary.

The concept of KNN is quite easy to understand. We first must have in our possession a training dataset where each point is labeled with the class it belongs to. We classify a new data point by finding the k training data points closest to it. Our new datapoint is then assigned to the class with that makes up the largest fraction of these k closest points. Intuitively, we are assigning class values based on the class with the highest (perceived) density in a given region.

An important consideration in the operation of a KNN classifier is the metric used to judge the similarity of the points. Minkowski distances of varying degrees are commonly used, the degree depending upon the application. The Minkowski distance is given by the formula:

$$D(a,b) = \left(\sum_{k=1}^{d} (a_k - b_k)^p \right)^{1/p} \tag{4.4}$$

where a_k and b_k are the values of the k^{th} feature of the points a and b respectively, p is the degree, and d is the dimensionality of the data points [20].

Because this distance must be calculated many times in order to classify a single datapoint, the computational complexity of KNN can pose a problem for real world applications. In a basic implementation, KNN is $\mathcal{O}(dn)$ for each classification, where d is the dimensionality of the dataset an n is the number of training samples. This is due to the fact that finding the k closest points to the sample being classified requires calculating the distance between the sample and each point in the training set. The computation performance can be improved by various data restructuring schemes, data pruning techniques, and calculation of partial distances.

Due of its simplicity, partial distance is an easy addition to the KNN algorithm. The partial distance is given by the formula

$$D_r(a,b) = \left(\sum_{k=1}^{r} (a_k - b_k)^p \right)^{1/p} \tag{4.5}$$

where r is the number of dimensions selected subject to the condition $r < d$, and p is the degree of the Minkowski distance [20]. If the partial distance is used outright, it will lead to inaccuracies since the distance in a subset of the dimensions is not representative of the full distance. However, we can take advantage of the fact that distance is monotonically increasing as we add more and more dimensional components. Thus we can still reach a correct distance comparison by comparing

a partial distance to another distance in the full space. If the partial distance already exceeds the full distance, then we may stop calculating. It is clear which point is farther away from the reference.

KNN has some convergence guarantees given an infinite number of training data points. It can be shown in the two-class case that an upper bound on the error rate is given by the smaller concave function greater than

$$\sum_{i=0}^{(k-1)/2} \binom{n}{k} \left[(P^*)^{i+1}(1-P^*)^{k-i} + (P^*)^{k-i}(1-P^*)^{i+1} \right] \tag{4.6}$$

where n is the number of training data points and P^* is the Bayes error rate. When $k = \infty$, the error rate for KNN converges to the Bayes error rate. Thus KNN is, in the limit, the same as the MAP classifier. Of course, in real life situations we do not have access to an infinite number of samples. In addition, we cannot choose k to be arbitrarily large. We desire for the k closest points to actually be close to the point we are evaluating so that there is some local influence over our decision. With a limited number of training points, we can only satisfy this locality objective by choosing a reasonably sized value for k [20].

As an aside, KNN can also provide insight into how high dimensionality can be problematic in classification. Consider that KNN is attempting, in the limit, to estimate the posterior distribution, $P(\omega_j|\boldsymbol{x})$. For a set number of data points, adding more features increases the "sparsity" of the dataset in the feature space, because every additional feature increases the volume covered by the distribution. Thus the locality of the posterior estimate is affected because the k nearest neighbors become further and further away in space. Stated another way, there may not be enough data to capture the trends of the probability density functions in higher dimensions [23].

We may also consider the fact that the contribution of any one directional component to the total distance is diminished. Imagine a dataset containing a few features whose values are strongly grouped by class. These features provide us with a large amount of useful information concerning the appropriate class label.

Now consider adding many more features which are poor indicators of class. The latter features, which do not give us much information, nevertheless contribute to the distance we use to determine the k points used for classification. The distance contributions from these uninformative features features can drown out the relative closeness of two points along some useful axis. In this way, the addition of more features may change the classification results, even when those features do not convey any useful information. It is here that dimensionality reduction techniques come into play. See Section 4.4.3 for more information.

4.4.1.3 Decision trees

An intuitive method of classifying a sample is to ask a series of questions about it, much like playing a game of 20 questions. This approach is useful because it does not require a geometric concept of distance between data points to exist. Without the need for a distance metric, we are free to include non-metric features, such as labels or colors.

We can use a decision tree to represent any particular series of questions. At the end of any given line of questioning lies a leaf node representing the class label we will assign to the datapoint in question. Importantly, it is possible to reform any decision tree into a binary tree which will yield the same answers [20]. This process involves converting the original questions into questions with simple yes/no answers.

So, given some labeled training data, we want to create a series of questions to determine the class label. However, one of the reasons we are using training data is that we do not understand the structure of the data. How then can we know which questions to pose? Fortunately it is possible to build a useful tree without advance knowledge if we consider our final goals. At the end of the growing period, we wish for similar data points to end up in the same leaf node. We also wish for the leaf nodes to represent only one of the class labels, so that the leaf nodes allow us to distinguish between classes. In this sense, we wish to maximize some metric of class purity at each leaf.

A general and commonly used tree growing procedure to accomplish these goals is CART (Classification And Regression Trees) [24]. In the case of a binary tree, we can think of each non-terminal node as representing a split based on some criterion. At each node, we choose to split the training dataset in such a way as to best separate the data, where "best" is determined by an evaluative function of our choosing. Typically, the best split is the one which most increases the class purity at the child node. Each split creates two child nodes; and the splitting rule is applied recursively on these children until some stopping criteria are met. An example stopping criterion could be complete purity at the leaf nodes [20].

As previously mentioned, the metric used to determine the best split is typically some notion of class purity. Instead of maximizing purity, it is more common to formulate the problem as a minimization of impurity. Two common impurity measures are the Gini impurity and the entropy. The Gini impurity for a node N is defined as:

$$i(N) = \sum_{i \neq j} P(\omega_i) P(\omega_j) = \frac{1}{2}[1 - \sum_j P^2(\omega_j)]. \tag{4.7}$$

It is the expected error rate at a given node if the class label is selected randomly from the classes present in N. The Gini impurity will have the highest value when the classes present have equal probability and will be zero when only one class is present. Similarly, the entropy impurity is defined as:

$$i(N) = -\sum_j P(\omega_j) \log_2 P(\omega_j). \tag{4.8}$$

This uses the notion of entropy from information theory. If all the points in a given node belong to the same class, the entropy is 0. The entropy finds its maximum value when each class is equally probable [20].

Although we can use the above equations to compare the effectiveness of splits, we have not specified the method by which the data is split. The simplest and most common method of generating splits is to consider thresholds on a single

feature. This type of splitting yields decision boundaries that are perpendicular to the feature axes. Leaf nodes generated in this way correspond to hyper-rectangular regions in the feature space. Though less common, the splitting function may be as arbitrarily complicated as desired. A small step up in complexity might be to calculate a hyperplane without restricting its directionality. Or we may choose to be very complicated and use another classification algorithm to construct a decision boundary at each node. Typically, simple methods perform well and are easier and faster to train [20].

4.4.1.4 Support vector machines

Support vector machines (SVM) are a type of binary classifier which seeks to separate samples using a hyperplane. Given two linearly separable classes, there are infinitely many hyperplane which will perfectly partition the data. The hyperplane selected by SVM is the one which has the maximum margin, i.e. the hyperplane which is the maximum distance from the closest samples [25]. These closest samples are called the "support vectors" and define the hyperplane. By choosing the hyperplane which allows for the most "wiggle room", it is hypothesized that we will be choosing the hyperplane which most generally applies to the data and is least subject to overfitting.

An SVM may be made into a nonlinear classifier by employing the kernel trick. The kernel trick allows any linear operation to be turned into a nonlinear operation by replacing the dot products with a nonlinear function $K(x, w)$. The kernel is used to perform a nonlinear mapping to a higher (even infinite!) dimensional space [26].

Any set of data can be made linearly separable by nonlinearly mapping to a new space with enough dimensions [20]. This fact motivates the use of high dimensional kernel functions. The kernel trick facilitates this by providing a computational shortcut. Without a shortcut, a high dimensional transformation would involve mapping two vectors into a high dimensional space, performing the dot product (an operation which is linear in the number of dimensions), and transforming the result back to the original space. Clearly when the number of dimensions is high, this is

computationally infeasible. The kernel trick allows for us to evaluate the result of this process directly without actually performing the mapping. One common example of a kernel function which demonstrates this is the Gaussian radial basis function kernel:

$$k(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2), \text{ for } \gamma > 0. \tag{4.9}$$

Using this kernel, the transformed features space can be shown to be a Hilbert space of infinite dimensions. However, as equation 4.9 shows, we may calculate this infinite dimensional dot product with a simple expression [20].

A concern when using infinite dimensional mappings is that we will have far too many free parameters in our decision boundary and will thus be subject to overfitting. It turns out, however, that because the hyperplane is defined by only the support vectors, the number of free parameters is exactly the number of support vectors, no matter the dimensionality of the transformed feature space. Due to this fact, support vector machines can provide a highly complicated decision boundary while still avoiding overfitting. As one might expect, however, the choice of the kernel will affect the classification performance of the SVM. Certain kernels will provide better results for some datasets [20].

4.4.1.5 Backpropagation neural networks

Neural networks are a machine learning construct loosely modeled on the way biological neurons are thought to work within the brain. A single artificial neuron consists of a set of weighted inputs and an activation which operates on the weighted sum of these inputs. Many artificial neurons can be connected together by allowing the output from one neuron be the input to one or many others. Some nodes in the network may be designated as input nodes and others as output nodes. It is into these nodes that the input data is pushed and from which the output signal is read, respectively. The exact topology of the network and the choice of activation function

will determine properties of the network. These topologies and their effects will not be discussed in detail here, as this information could and has filled many books and is an ongoing subject of research in the machine learning community [20].

The activation function in an artificial neuron is usually chosen to be a nonlinear, smoothly differentiable function. A nonlinear activation function is desirable because it allows for nonlinear decision boundaries to be described. Differentiability is desired because it allows for training using backpropagation, a technique which will be discussed shortly [27]. A neural network is typically trained using a large set of input pattern and corresponding desired output patterns. In this scheme, each input pattern is presented to the network and the weights are subsequently tuned such that the network output is as close to the desired output pattern as possible. The output pattens can take a variety of forms. In supervised classification, the output will represent some class label associated with the input pattern. In the unsupervised case, such as in a feature detection task, the network may simply be asked find some internal representation which allows it to reproduce the input [28][29][30].

How can a network be trained to produce the output that we desire? This is a question with an interesting history. In short, however, the answer is to use backward propagation of errors, or backpropagation [27]. This is an method for calculating the gradient of some loss function in terms of the individual weights within a network. As it turns out, a closed form expression for the partial derivative of the error with respect to each weight is obtainable. When paired with an optimization method like gradient descent, the weights can be incrementally updated to find a configuration which minimizes a loss function, e.g. mean squared error, over the training data.

4.4.2 Ensemble methods

4.4.2.1 Naive Bayes

Naive Bayes is a fusion technique that assumes the predictions of several classifiers based on a given dataset are independent. Given some probability distribution for

the true class ω given the label predicted by the j^{th} classifier $\hat{\omega}_j$ and our independence assumption, we may write the probability for the true class as

$$P(\omega \mid \hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_N) = \prod_{j=1}^{N} P(\omega \mid \hat{\omega}_j) \tag{4.10}$$

We can choose a label for a given sample that maximizes this probability. The individual probability distributions may be obtained by applying a test set to the classifier and creating a confusion matrix. Normalizing the rows for each output class $\hat{\omega}$ gives a probability distribution over the true class values ω. A score for each label can be given by the probability distribution [20][31].

4.4.2.2 Majority Vote

Majority vote is a conceptually simple fusion method that simply chooses the label with the most votes from the ensemble. Explicit upper and lower bounds exist for the combined accuracy when majority voting is employed. It is also possible to calculate the probability that an ensemble will have an accuracy higher than the greatest individual classifier accuracy. More information on this may be found in reference [32]. In order to determine a label in a majority voting scheme, we will classify a given sample with each classifier and then choose the sample with the most votes, breaking ties randomly. A score can be given by the fraction of votes assigned to each class.

4.4.2.3 Behavior knowledge space

Behavior Knowledge Space (BKS) is a fusion technique that builds a lookup table of the most probable class given the output of the classifiers in the ensemble[33]. The chosen class need not be the class suggested by the ensemble. A BKS lookup table is constructed by classifying the points in a training set and creating a map from all observed combinations of outputs from the classifiers in the ensemble. The

key in this map is the list of classifier outputs, while the value is a histogram of the true class value when the ensemble produced the given outputs. Once the map has been formed, a subsequent point may be classified by finding the output from the ensemble, looking up the histogram, and choosing the most probable label. A score may be given by normalizing the histogram. One potential problem with BKS occurs when the number of possible class labels increases. With a fixed number of training samples, the lookup table will become more and more sparse, causing the probability of a miss happening when classifying a new sample. In the implementation used here, a majority vote was used to handle these misses.

4.4.2.4 Random forest

Random Forest (RF) is an ensemble classifier, meaning it combines the results of other sub-classifiers to form a final answer. In the case of RF, the sub-classifiers are decision trees, which have been grown with some degree of randomness. This randomness is what makes RF such a competitive classifier. It has been demonstrated that increasing the level of randomness to extreme degrees can actually yield good results for many datasets [34]. To a large degree, the utility of the random component is in avoiding over-fitting. Random forest can also be shown to employ a neighborhood approach somewhat like kNN [35].

Random forest was originally presented by Breiman [36]. Although Breiman had a specific methodology for generating his trees, the idea of RF is actually very flexible. Several design choices can be made and implemented at various stages of the algorithm. These design choices will give the forest different properties to help tailor it to the specific problem. In that sense, RF can be thought of as more of a concept than a specific algorithm. While most of the design choices involve options like those suggested in the previous discussion of CART, others, like the method of injecting randomness, are specific to RF. Of course, great success with the original implementation is what generated academic interest in RF, so alteration is not necessarily needed to achieve good results.

Breiman's implementation for RF was as follows. Given N data points with M features, grow a forest of n_{tree} trees. Train each tree using a subset of the data drawn with replacement from N. Grow the tree using a procedure like CART, except that only a random subset, m_{try}, of the M features is evaluated for splitting at each node. The feature within m_{try} that best separates the data is used to generate the split for the node. The trees are fully grown (that is, grown until no more splits can be made) and are not pruned. Once the forest is grown, a new data point is classified by allowing each individual tree to categorize it. A plurality vote amongst the individual trees is the basis for the final classification. Thus, random forest is a type of majority vote ensemble (Section 4.4.2.2).

4.4.3 Dimensionality reduction

Dimensionality reduction is an important tool when tasks suffer from the curse of dimensionality. In machine learning problems, the curse of dimensionality is related to the increasing sparsity of a dataset in the problem space as the dimensionality of the dataset increases. In short, more and more data is required to adequately represent the high dimensional structure of a dataset's underlying distribution. Computational and storage costs also increase when working in higher dimensions. In many cases it is not feasible to acquire, store or process enough data to reach a solution.

Oftentimes, however, the information in the original dimensions is not as dense as it could be. When features in the dataset are correlated with one another, the information in each individual feature is reduced. In these cases, we can use dimensionality reduction techniques to find a new set of derivative features which represent the same information in a smaller space. If we can reduce the number of dimensions sufficiently without introducing too much information loss, our problem can become tractable without drastically reducing the accuracy of our solution.

In this section, we will explore two dimensionality reduction methods with differing end goals. Principal component analysis (PCA) seeks to find the best representation

of the data possible. It essentially finds a rotation of coordinate system where the basis vectors follow the direction of maximum variance. Dimensionality reduction is accomplished by ignoring the directions along which the variance (and thus information) is minimal [37]. Fisher's linear discriminate (FLD), on the other hand, is designed with classification accuracy in mind. It finds a way to project the data such that the transformed class distributions are best separated from one another, regardless of how much this obscures the structure of the original distributions [38].

In addition to performing dimensionality reduction, these techniques have seen extensive use in spectral anomaly detection and radioisotope identification (see [14] [18] [19] [39] [40]).

4.4.3.1 Principal component analysis

Please note that the derivations in this and the following section are borrowed heavily from [20].

If we have a set of d-dimensional samples, we might be interested to find a 1-dimensional representation of the data such that the sum of the squared distance between the projection direction and the samples is minimized. In other words, if we want to reduce our set of samples from d dimensions to one dimension, which axis should we choose to minimize the loss of information? Let the equation of the axis line be written as:

$$\boldsymbol{x} = \boldsymbol{m} + a_k \boldsymbol{e} \tag{4.11}$$

where $\boldsymbol{m}$ is the mean of the distribution, $\boldsymbol{e}$ is a unit vector in the direction of the line, and a_k is a scalar representing the distance of the sample point x_k from the line. If we wish to find a line that minimized the squared error, we can solve for the optimal set of coefficients by minimizing:

$$J(a_1, ..., a_n, \boldsymbol{e}) = \sum_{k=1}^{n} ||(\boldsymbol{m} + a_k \boldsymbol{e}) - \boldsymbol{x}_k||^2 \tag{4.12}$$

If we take the derivative of this with respect to a_k and set the results equal to zero, we are left with:

$$a_k = \boldsymbol{e}^\top(\boldsymbol{x}_k - \boldsymbol{m}) \tag{4.13}$$

Plugging this result back into equation 4.12 yields

$$J(\boldsymbol{e}) = -\boldsymbol{e}^\top \boldsymbol{S}\boldsymbol{e} + \sum_{k=1}^{n} ||\boldsymbol{x}_k - \boldsymbol{m}||^2 \tag{4.14}$$

where $\boldsymbol{S}$ is the scatter matrix, defined as

$$\boldsymbol{S} = \sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^\top \tag{4.15}$$

To minimize $J(\boldsymbol{e})$ we must maximize $\boldsymbol{e}^\top \boldsymbol{S}\boldsymbol{e}$. Through some manipulation it can be shown that $\boldsymbol{e}$ must be an eigenvector of the scatter matrix $\boldsymbol{S}$:

$$\boldsymbol{S}\boldsymbol{e} = \lambda e. \tag{4.16}$$

Specifically, $\boldsymbol{e}$ must be the eigenvector with the largest eigenvalue. We can extend our 1-dimensional representation to a d'-dimensional representation by projecting on the first d' eigenvectors when they are arranged in descending order of their eigenvalues.

4.4.3.2 Fisher's linear discriminant

It is not always the case that minimizing reconstruction error is our main goal. We may instead seek to maximize the separation of the projected class distributions. Let us consider the two class case. Let $\widetilde{m}_i$ and $\widetilde{s}_i$ represent the mean and scatter of the i^{th} projected class respectively. If $\boldsymbol{w}$ is the projection direction, then we seek to maximize

$$J(\boldsymbol{w}) = \frac{|\widetilde{m}_1 - \widetilde{m}_2|^2}{\widetilde{s}_1^2 + \widetilde{s}_2^2}. \tag{4.17}$$

If we define the class i scatter matrix $\boldsymbol{S}_i$ as in equation 4.15 and $\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2$, then we can write

$$\widetilde{s}_i^2 = \sum_{\boldsymbol{x} \in \mathcal{D}} (\boldsymbol{w}^\top \boldsymbol{x} - \boldsymbol{w}^\top \boldsymbol{m}_i)^2 = \boldsymbol{w}^\top \boldsymbol{S}_i \boldsymbol{w}. \tag{4.18}$$

Thus

$$\widetilde{s}_1^2 + \widetilde{s}_2^2 = \boldsymbol{w}^\top \boldsymbol{S}_W \boldsymbol{w}. \tag{4.19}$$

The term involving the projected means has a similar form:

$$(\widetilde{m}_1 - \widetilde{m}_2)^2 = \boldsymbol{w}^\top \boldsymbol{S}_B \boldsymbol{w}. \tag{4.20}$$

where

$$\boldsymbol{S}_B = (\widetilde{m}_1 - \widetilde{m}_2)(\widetilde{m}_1 - \widetilde{m}_2)^\top. \tag{4.21}$$

Substituting equations 4.19 and 4.20 into equation 4.17 yields

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^\top \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{S}_W \boldsymbol{w}} \tag{4.22}$$

The vector $\boldsymbol{w}$ that maximizes this equation must also satisfy

$$\boldsymbol{S}_W^{-1} \boldsymbol{S}_B \boldsymbol{w} = \lambda \boldsymbol{w}. \tag{4.23}$$

However, we know that $\boldsymbol{S}_B \boldsymbol{w}$ is always in the direction of $\boldsymbol{m}_1 - \boldsymbol{m}_2$, so we need not solve the eigenvalue problem. Instead we can simply write

$$\boldsymbol{w} = \boldsymbol{S}_W (\boldsymbol{m}_1 - \boldsymbol{m}_2). \tag{4.24}$$

4.5 1-D Imaging Study

In this study, we sought to understand the benefit that supervised classification algorithms could have on threat source detection in coded aperture imagery. For

machine learning to be useful, it would need to outperform the simple threshold alarm methods we had been using in the past. The ability to fuse neutron and gamma ray data were also of interest.

The threat sources modeled were a gamma and neutron emitting threat source in various shielding configurations. Gamma ray background was modeled as being emitted from a thin cylinder of concrete below the dual imager, as most gamma ray background at the energies of interest is expected to come from the ground. The gamma ray background was varied every 5 meters by sampling from a distribution of cross country intensities as measured by the Large Area Imager (LAI). The neutron background was emitted from a sphere around the system with no changes in intensity and an energy distribution as suggested by Goldhagen [41]. Coded aperture images were reconstructed in post processing of the raw detector data by projecting the mask pattern back onto the scene from each simulated interaction position. An example of a 1d coded aperture image reconstructed at a single distance from simulations is shown in figure 4.1. Features were extracted from these images and used to train a variety of supervised classification algorithms. The performance of each algorithm was evaluated using 10-fold cross validation. Additionally, the effect of dimensionality reduction and combination of classifiers into ensembles was also investigated.

It should be stressed that the data which will be fed into the final alarming algorithms developed for the DDLI may be significantly different than the dataset used in this investgation. Despite this, the dataset provides some common and project-related ground on which to compare different alarming mechanisms. Thus the numeric values presented in the investigation should be treated as indicators of relative algorithm performance and not as representative of the concrete alarming capabilities of the DDLI system.

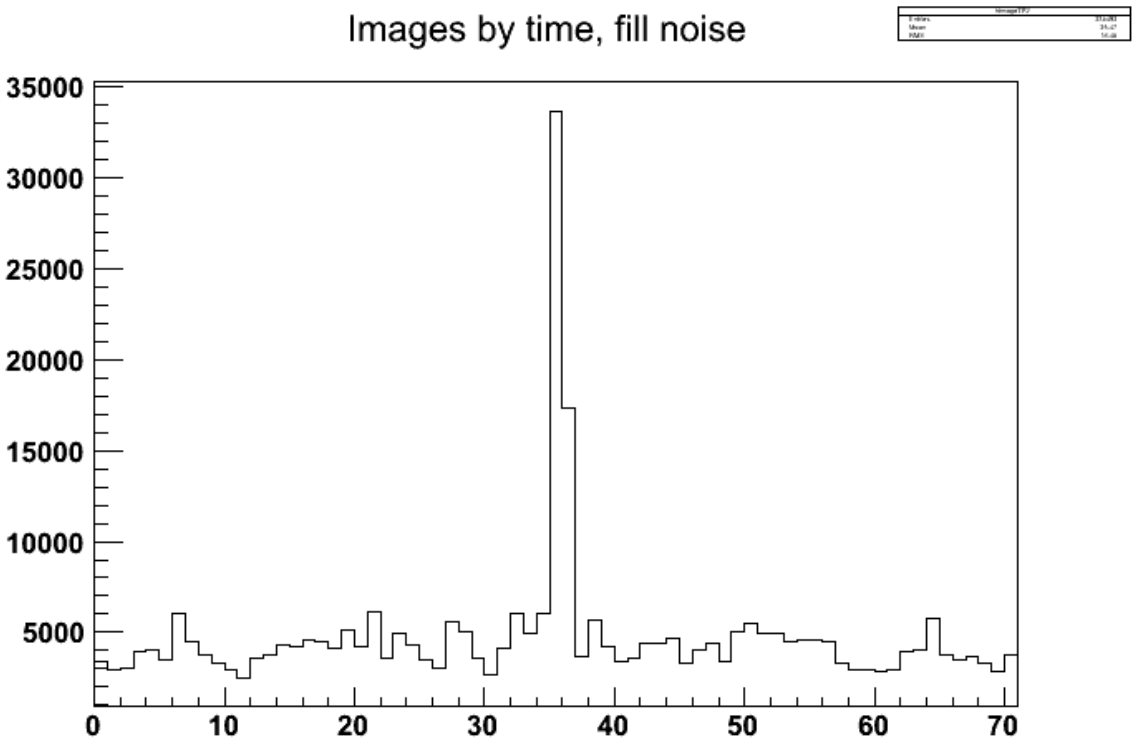


Figure 4.1: Image of a point source located at the 35 meters mark along a simulated driving path. This is an image reconstructed at the correct source distance.

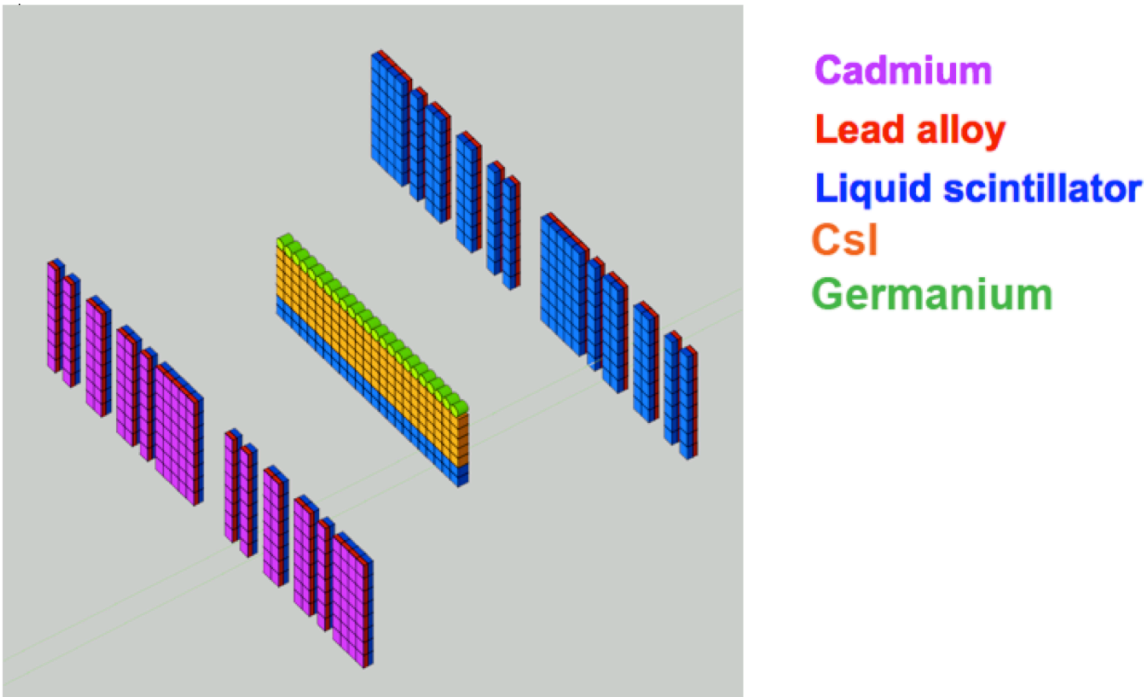


Figure 4.2: Structural diagram of dual imager detector systems.

4.5.1 Dataset

The dataset consists of 1594 samples from 16 classes. Each sample in the original dataset consisted of 28 features. The 28 features can be broken into four sets of seven features. The same seven features were extracted from each of the four coded aperture images originating from the CsI, HPGe, l liquid scintillator gamma and liquid scintillator neutron detector arrays. The seven features extracted from each array image were the maximum pixel value, the values of the two pixels neighboring the maximum pixel, the image integral, the mean and standard deviation of the image pixels, and the signal to noise ratio of the maximum pixel. The 16 classes correspond to both background and source plus background for 8 different source, imaging-distance, and speed configurations. These source configurations are specifically 20, 30, and 50 meter standoff, shielded with 1cm of tungsten or unshielded, and 5 or 10 m/s driving speed. This dataset was normalized to have zero mean and unit variance in all features.

A version of this dataset was generated by applying principal component analysis. The dataset was reduced to 9 features while maintaining greater than 99% reconstruction accuracy as determined by the PCA eigenvalues. This dataset was renormalized to zero mean and unit variance in each feature.

Another version of the data was produced using Fisher's linear discriminant. In this case, FLD was applied 16 times in a one vs. all manner to produce a new space of projection directions. Thus the dimension was reduced to 16 dimensions. This was reduced further in a separate dataset to 11 dimensions with over 99% reconstruction accuracy by applying PCA. As before, these datasets were normalized to zero mean and unit variance in each feature.

Because the problem may be condensed to a two-class case by considering only a positive case where the source was present and a negative case where the image contains only background, the performance of the classifiers on a separate binary labeling scheme was also evaluated.

4.5.2 Gaussian maximum posterior probability

The parametric maximum posterior probability classifiers tested in this investigation were based on Gaussian fits to the data. Three different cases were examined. In the first case, it was assumed in the fit that each distribution had no correlation in the features. Furthermore, it was assumed that each class had an identical covariance matrix. This matrix was calculated by finding the average covariance matrix across all the classes and removing the covariance terms. In the second case, it was assumed that all the classes were distributed according to the same covariance matrix, but covariance terms were not removed. In the final case, each class was fitted with a Gaussian based on its own unique covariance matrix. The class was predicted for any new sample by choosing the class which maximized the posterior probability according to the fitted gaussian distribution and any supplied priors. The scores for each possible label were given by the posteriors for each class.

Table 4.1: Accuracy of MPP classifiers on complete label set based on 10-fold cross validation

	Full	FLD	PCA
MPP case 1	0.6218	0.6587	0.6330
σ	0.0214	0.0283	0.0343
MPP case 2	0.6976	0.7096	0.6474
σ	0.0210	0.0366	0.0237
MPP case 3	0.6267	0.6663	0.6487
σ	0.0193	0.0304	0.0113

Table 4.2: Accuracy of MPP classifiers on binary label set based on 10-fold cross validation

	Full	FLD	PCA
MPP case 1	0.6920	0.7767	0.7697
σ	0.0486	0.0277	0.0300
MPP case 2	0.8175	0.8244	0.7679
σ	0.0379	0.0274	0.0310
MPP case 3	0.8087	0.8231	0.7961
σ	0.0271	0.0254	0.0208

4.5.3 K-nearest neighbor

K-Nearest Neighbor classifiers were employed using a variety of values for the neighborhood parameter k. The kNN classifier was generated using MATLAB's

ClassificationKNN class. The class was predicted for a sample point by choosing label accounting for a plurality of the k nearest points in the training set. The scores for each label were given by the fraction of the k neighbors represented by each class.

Table 4.3: Accuracy of the KNN classifier on both full and binary label sets based on 10-fold cross validation

	Full	FLD	PCA
KNN full labels	0.6199	0.6581	0.4573
σ	0.039	0.035	0.0275
KNN binary labels	0.7717	0.8214	0.7183
σ	0.0305	0.0424	0.0211

4.5.4 Decision tree

Decision trees were grown using MATLAB's classregtree function. Gini impurity was chosen as the split criterion and the tree was pruned to its optimal size as determined by a test set. The predicted class for a sample was chosen by taking the plurality label in the terminal leaf node. The score for each label was given by the fraction of training samples accounted for by each class in the leaf node.

Table 4.4: Accuracy of the Decision Tree classifier on both full and binary label sets based on 10-fold cross validation

	Full	FLD	PCA
DTree full labels	0.6669	0.628	0.6016
σ	0.0307	0.0281	0.0302
DTree binary labels	0.8268	0.7924	0.7772
σ	0.0301	0.0347	0.037

4.5.5 Support vector machine

Support vector machines were trained using both linear decision boundaries and the Gaussian Radial Basis Function kernel. Appropriate parameters for the RBF kernel were determined via cross validated optimization. Because support vector machines are inherently binary classifiers, one SVM was trained for each class using a one vs. all approach. The predicted label for a sample was chosen by a random split among classes who claimed it belonged to their class. The score was given by a constant value for each “claiming" SVM equal to one over the number of candidate classes and zero if the SVM did not claim the sample.

Table 4.5: Accuracy of the SVM classifier on both full and binary label sets based on 10-fold cross validation

	Full	FLD	PCA
SVM full labels	0.6562	0.6487	0.5038
σ	0.0147	0.0273	0.0202
SVM binary labels	0.8419	0.8494	0.8131
σ	0.0323	0.0219	0.0217

4.5.6 Backpropagation neural network

A back-propagation neural network was trained using MATLAB's patternnet function. This produces a neural network with a specified number of units and hidden layers. The hidden layers use a sigmoid activation function and connections between each layer are complete. Trial and error was used to come up with a network topology consisting of two hidden layers with 16 nodes each. A label was assigned to a sample by choosing the most activated output node (one for each class). The score was given by the normalized activations of each output node.

Table 4.6: Accuracy of the BPNN classifier on both full and binary label sets based on 10-fold cross validation

	Full	FLD	PCA
BPNN full labels	0.7039	0.7127	0.6625
σ	0.0352	0.0268	0.0309
BPNN binary labels	0.8325	0.8494	0.8149
σ	0.0368	0.0188	0.0271

4.5.7 Discussion of single classifier performance

The confusion matrices for the various classifiers indicate that most classifiers had trouble distinguishing background classes from one another, which is to be expected. In terms of source plus background classes, 10, 13, 15, and 16 were the most prone to misclassification, usually being mistaken for background and not other source types. These correspond to the shielded and long distance/high speed cases.The BPNN performed best and most consistently with the fully labeled data, followed by the case 2 Gaussian MPP. In the binary labeled case, BPNN was once again a consistent performer. The accuracy of SVM also improved.

The FLD dimensionality reduction generally improved the performance of the individual classifiers with the exception of decision trees in both the full and binary label cases. PCA reduction was in general detrimental to classification accuracy, sometimes dramatically so.

4.5.8 Application of ensemble classification methods

Ensemble classifiers aggregate devices which combine the output of other classifiers to produce a new (and hopefully better informed) result. Random Forest, an ensemble classifier, has been previously applied to certain aspects of this imaging dataset with good results. For this effort, three different methods, Naive Bayes, Majority Vote, and Belief Knowledge Space were investigated.

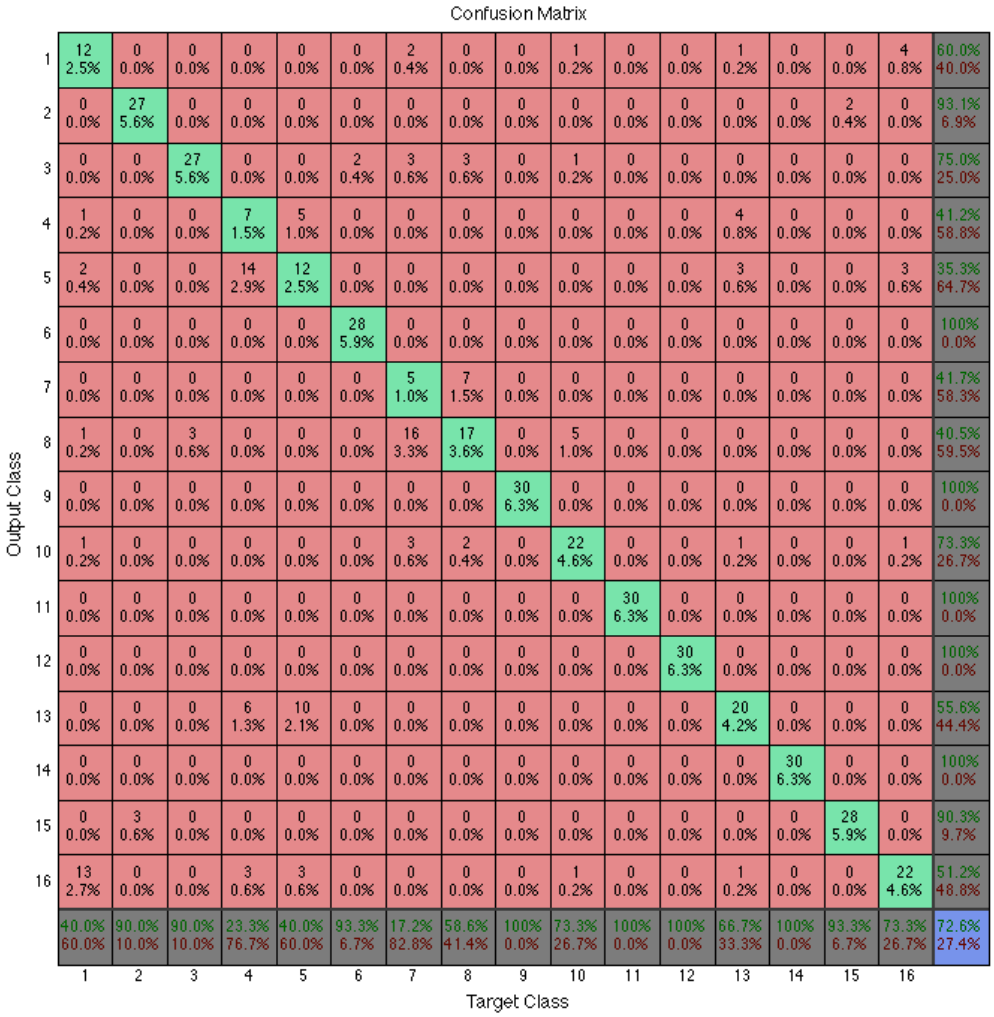


Figure 4.3: Confusion matrix for the BPNN classifier on the full dataset with full labels.

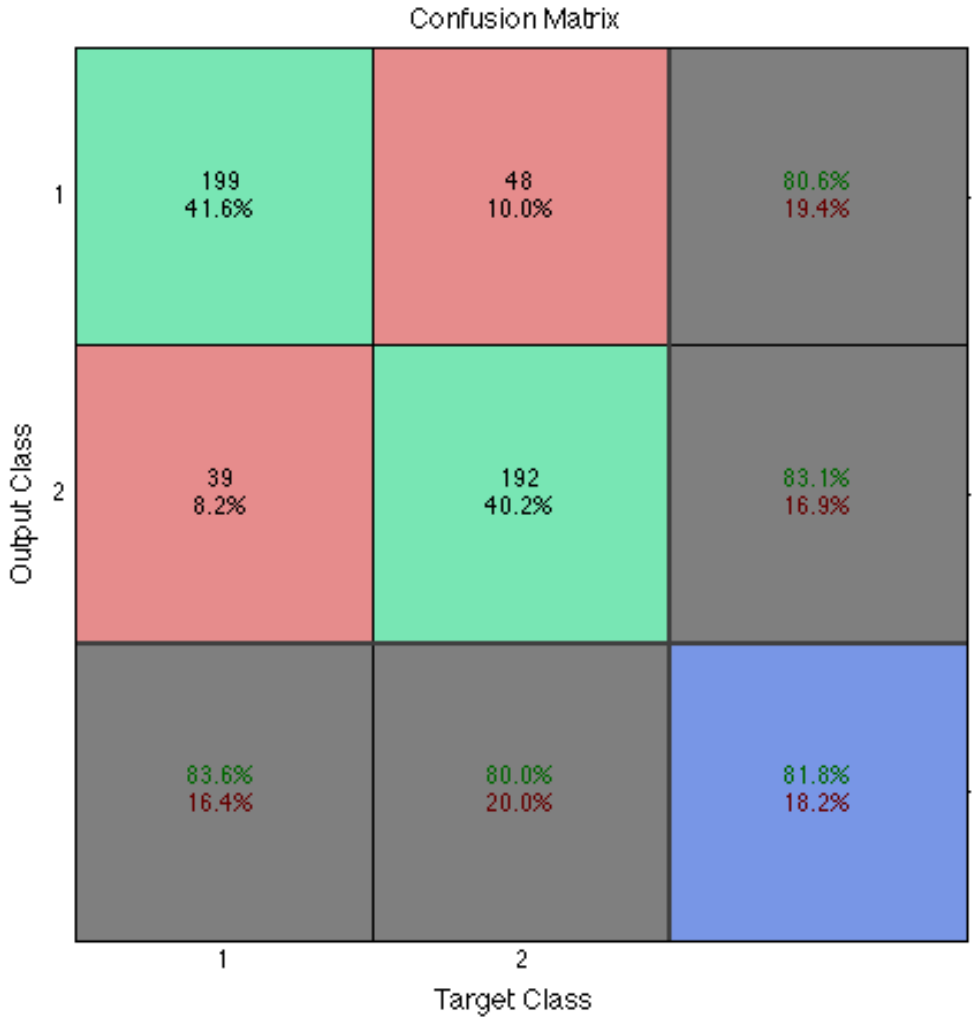


Figure 4.4: Confusion matrix for the BPNN classifier on the full dataset with binary labels.

Table 4.7: Accuracy of the fused and individual classifiers. This is a group of the best performing individual classifiers. The performance was evaluated on the full featured dataset with full labels using 10-fold cross validation.

GOOD GROUP				
	ensemble	MPP2	BPNN	DTree
Naive Bayes	0.6951	0.7026	0.6789	0.6487
σ	0.0572	0.0361	0.0466	0.0334
Majority Vote	0.7001	0.6788	0.6675	0.6762
σ	0.0486	0.0347	0.0342	0.0345
BKS	0.7113	0.6838	0.6763	0.6863
σ	0.0479	0.0273	0.0467	0.0325

Table 4.8: Accuracy of the fused and individual classifiers. This is a group of the best performing individual classifiers. The performance was evaluated on the full featured dataset with binary labels using 10-fold cross validation.

GOOD GROUP BINARY				
	ensemble	MPP2	BPNN	DTree
Naive Bayes	0.8281	0.8143	0.8231	0.8219
σ	0.0603	0.0445	0.0348	0.0322
Majority Vote	0.8382	0.8194	0.8395	0.8294
σ	0.0324	0.0227	0.0374	0.0373
BKS	0.8356	0.8242	0.8181	0.7993
σ	0.0298	0.0291	0.0391	0.0493

Table 4.9: Accuracy of the fused and individual classifiers. This is a mixed group with two poor classifiers and one good classifiers. The performance was evaluated on the full featured dataset with full labels using 10-fold cross validation.

MIXED GROUP				
	ensemble	MPP2	BPNN	DTree
Naive Bayes	0.7126	0.6425	0.6675	0.6111
σ	0.0375	0.0337	0.0342	0.0405
Majority Vote	0.6537	0.6424	0.6813	0.6024
σ	0.026	0.0215	0.0366	0.049
BKS	0.7090	0.6424	0.6849	0.6048
σ	0.0263	0.0350	0.0574	0.0472

Table 4.10: Accuracy of the fused and individual classifiers. This is a mixed group with two poor classifiers and one good classifiers. The performance was evaluated on the full featured dataset with binary labels using 10-fold cross validation.

MIXED GROUP BINARY				
	ensemble	MPP2	BPNN	DTree
Naive Bayes	0.8167	0.7100	0.8231	0.7188
σ	0.065	0.0674	0.0348	0.0531
Majority Vote	0.718	0.6852	0.8168	0.7303
σ	0.044	0.0602	0.0307	0.0401
BKS	0.8331	0.7100	0.8231	0.7188
σ	0.0527	0.0674	0.0348	0.0531

Table 4.11: Accuracy of the fused and individual classifiers. This combination of the two individual classifiers with the best performance yielded the highest ensemble accuracy. The performance was evaluated on the FLD dataset with binary labels using 10-fold cross validation. Multiple trainings were required to get these results.

BEST GROUP BINARY			
	ensemble	SVM	BPNN
BKS FLD	0.8633	0.8581	0.8531
σ	0.0373	0.0343	0.0460

4.5.9 Discussion of ensemble performance

Using ensemble methods generally resulted in performance at least comparable to the best classifier in the group. Majority Vote seemed to be most susceptible large fluctuations in performance gain. This is due to the fact that each classifier in the ensemble plays an equally important role in deciding the label for a given sample. As such, the general, as opposed to region specific accuracy of each classifier is more directly responsible for ensemble performance. Majority vote also does not take the confidence of each classifier into account. This effect can be seen in the mixed trials, where majority vote often achieves some middle ground value for accuracy. Majority vote might possibly be made more resilient if the score for each classifier was combined instead of their single vote.

Unlike majority vote, BKS and Naive Bayes were more resilient to having a mixture of good and bad classifiers in the ensemble. They were generally able to maintain a level close to or significantly above the accuracy of the best classifier. Both of these methods could likely have benefited from more ensemble training data, and all ensemble methods may have benefited from the injection of some randomness,

such as having their classifiers trained on different subsets of the data. It has been shown that randomness can improve generalization qualities in ensemble classifiers. [34].

4.5.10 Conclusions

It has been demonstrated that certain single classifiers such as BPNN perform somewhat better on this dataset than others. It has also been observed that dimensionality reduction with FLD can improve the performance of classifiers, while PCA dimensionality reduction can have the opposite effect. Ensemble classifiers were not able to exhibit drastically increased performance over the best individual classifiers on this dataset, but were able to yield a few percentage points of improvement. Ensemble methods were, however, able to give a much improved answer over their individual members when the member performance was low.

The performance of many supervised classification algorithms have now been evaluated and compared on a difficult mixed dataset of simulated 1D coded aperture imagery from the optimized design. While this dataset was not exactly representative of the data which will be processed by the final system, it did indicate the relative strength of the algorithms on data of this nature. It also gave us the opportunity to compare machine learning techniques against the single feature thresholds which were previously employed. A comparison of ROC curves for two single feature thresholds and a BKS ensemble of supervised classifiers can be seen in figure 4.5. Backpropagation Neural Networks and Non-linear Support Vector Machines were found to give good performance and generally outperform the single feature thresholds. In addition ensembles were shown to be able to significantly improve results when the individual performances of the member classifiers was not as good as possible.

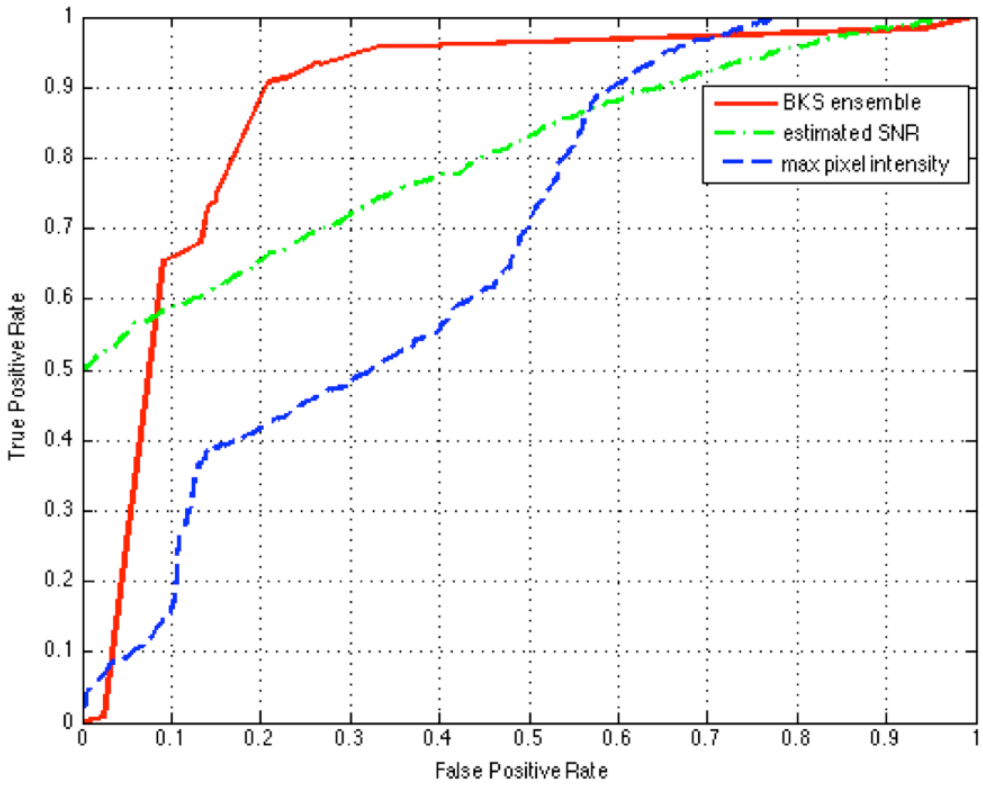


Figure 4.5: Performance comparison between an ensemble of classifiers and two single metric criteria on a difficult dataset. ROC curves are shown for a BKS ensemble, the estimated SNR, and the max pixel intensity. Data was simulated based on the Dual DLI geometry.

4.6 LAI Study

4.6.1 Overview

A major concern in simulations of mobile imaging systems is accurately representing background. The background signal measured by a moving system contains high variability and a structure related to the environment. These qualities make it difficult to reproduce. In order to have a realistic platform for algorithm development and confidence in the reported results, it was decided to incorporate real measured background where available. The Large Area Imager (LAI) dataset, discussed in Section 4.6.2, makes this type of measured data available. In this section, a source injection study using the LAI cross country dataset will be described.

4.6.2 Large Area Imager dataset

In 2006, Large Area Imager (LAI, Section 5.2.1.1) was used to take a cross country background dataset. This data was acquired over the course of nine days by driving the imager from Livermore, California to Oak Ridge, Tennessee. During the journey, the imager passed through several other cities including Phoenix, AZ, Amarillo, TX, and Little Rock, AR.

The data stream from the imager was recorded in binary files using a custom format. These binary files contain ADC (Analog to Digital Converter) readings and an ID for each detector hit, counts from a "fifth wheel" representing distance traveled, and periodic GPS information.

Although the large area imager was not equipped with neutron detectors, this gamma-only background dataset was still useful for developing feature extraction and alarm algorithms. This is due to the fact that coded aperture images should appear similar for a point source regardless of the particle being measured. Similarly, the classification algorithms being investigated could easily incorporate more data from neutron sensing modalities.

One challenge in using the large area imager dataset was that noticeable energy drift occurred during the cross country run. Additionally, ADC channels were not converted into energy units before being saved to the binary files. Thus, an energy calibration would be required in post processing before using the measured spectra in a source injection study. Pure counts, however, are not affected by energy calibrations. For this reason, the coded aperture images generated from the LAI background dataset for this study used all counts with no spectral cuts.

4.6.3 Simulations

In order to conduct the source injection study, it was necessary to produce a dataset of simulated source passes to be injected. As a prerequisite for these simulations, a Geant4 model of the Large Area Imager needed to be created. Using information from an early SolidWorks model of the imager in addition to measurements taken by hand, the model shown in 4.6 was constructed. The model included 54 CsI detectors and dual lead lined masks. In the interest of time, no framing or truck geometry was recreated. While these components should have some effect, they would mostly serve to advantageously reduce background radiation coming from above and below the imager which would not be modulated by the mask. The model was raised above the ground to match the height of the LAI trailer. The ground itself was simulated as a very large concrete slab extending in all directions.

To be consistent with earlier simulations (Section 5.3), an existing weapons grade plutonium source model was used. Each simulated run consisted of a 600 meter pass by the source. More than 10,000 source passes were simulated with standoff distances ranging from 30-70 meters at driving speeds of either 5 or 10 meters per second. No obstructions other than the ground were simulated.

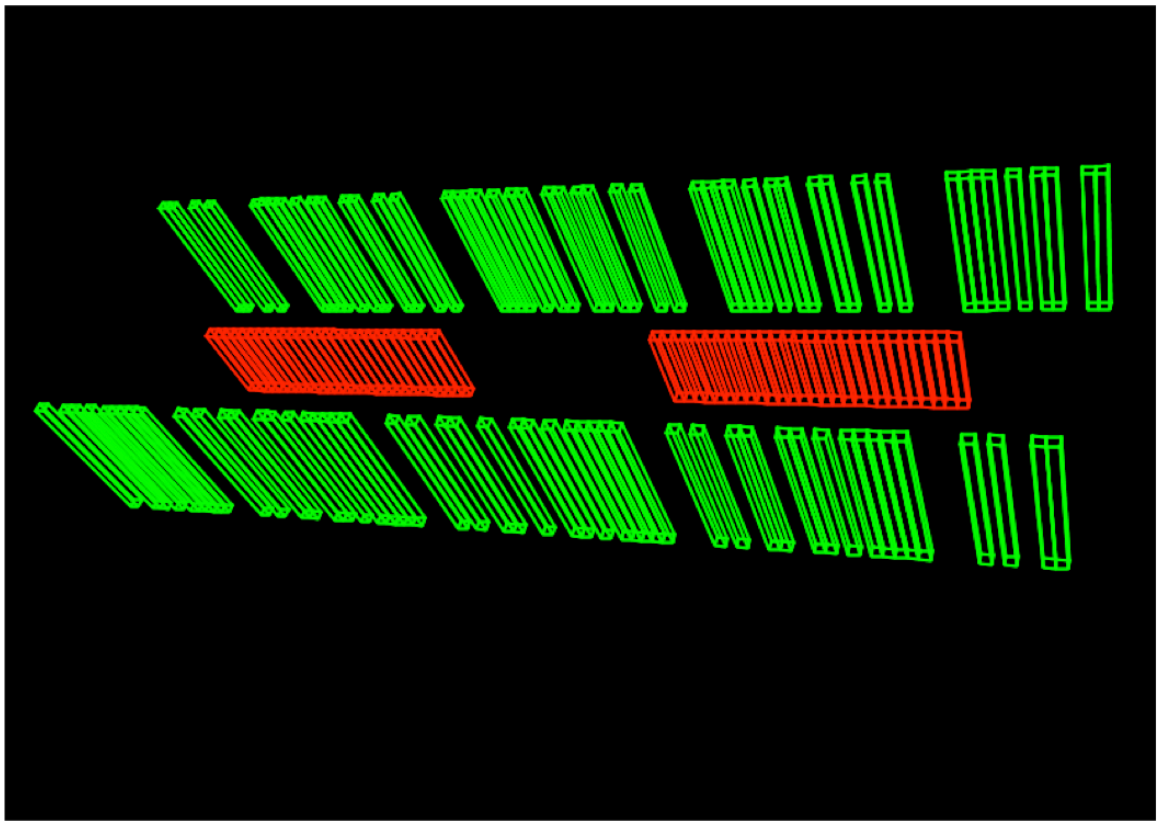


Figure 4.6: A wireframe model of the large area imager geometry. The CsI detector array is shown in red while the lead mask elements are shown in green.

4.6.4 Reconstruction

In order to match the fixed speed simulations with the variable speed LAI data, reconstruction of the measured background data was carried out under a very important assumption. Because the intensity of the gamma ray background is fairly stable over short distances, the same number of gamma rays should be recorded by the LAI during a given time frame regardless of its speed. Thus it was decided to time sample the measured LAI background and calculate the world-position of each gamma event based on the speed at which the injected source was simulated. This is as opposed to using the navigation data recorded by the LAI to place the gamma event in the world. In short, the LAI data, which represents real world driving at a range of speeds, was reconstructed as if it was taken at a single, constant speed.

Using this assumption, it is possible to report the performance of the alarm algorithms at set standoff distances and speeds. This would not be possible using the recorded speeds and scaling the features to account for dwell time. Despite this deviation from reality, the time-sampled LAI background is expected to allow for fairly realistic source injection. As most of the LAI data was taken at highway speeds, the typical operation will be a slowing of speed. This will have the effect of spatially compressing background variations. This compression should increase the variance within individual images and could make detection more difficult.

To perform the source injection, data from the Large Area Imager and the simulated source passes were reconstructed separately using the data analysis framework detailed in Section A.2. Within this software, modules were created to unpack both types of data into a common format, handle the fixed-speed navigation, coded aperture reconstruction, proximity localization reconstruction, and the saving of images to disk.

Each file from the LAI cross country data was effectively concatenated together into a single long file and processed as a continuous data stream. This data was streamed in the order it was taken to preserve background continuity. Coded aperture

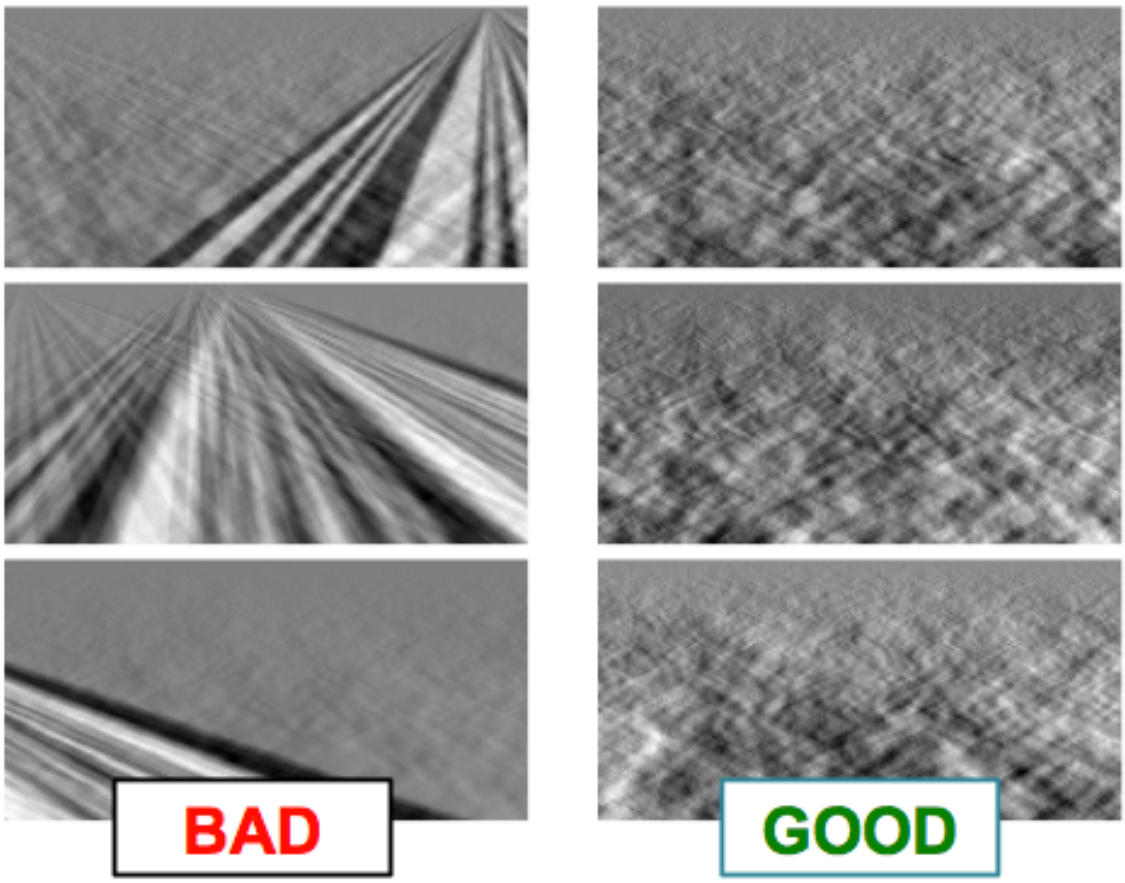


Figure 4.7: Examples of artifacts caused by the bad detector channel. The left column show background-only scenes reconstructed with the bad channel included. The right column shows the same scenes with the bad channel disabled.

and proximity localization images were generated for every 200 meters of travel and covered 100 meters of standoff. Simulated images were processed in a similar fashion and quantized to prevent an overlap.

During the reconstruction procedure, it was noticed that many of the LAI background images were very poor in quality and showing noticeable artifacts. It was discovered that one detector channel in the LAI system was bad and frequently reported bursts of spurious events. When this channel was disabled, background reconstructions were greatly improved. For this reason, the signal from one LAI detector was omitted from both source and background coded aperture reconstructions. Examples of these artifacts can be seen in figure 4.7.

4.6.5 Analysis

In the first stages of this analysis only the 10 m/s coded aperture imaging data was analyzed. This corresponded to a dataset of 5000 source images, 1000 each from 30, 40, 50, and 60 meter standoff distances. A total of 5000 background images remapped to 10 m/s were also used. Two feature sets were extracted for the purposes of this analysis. One feature set was small, containing only 8 elements. The second was larger, with 79 total features. The details of these feature sets are listed in Figure 4.6.5. A Random Forest classifier was trained separately on each standoff distance. Out of bag (OOB) error estimates (see Section 4.4.2.4) were used for early reporting of performance.

Small feature set	Large feature set
• Max pixel value	• Max pixel value
• SNR	• Count of MSER regions
• Min pixel value	• Count of BRISK points
• Mean pixel value	• Statistics for 5x5 region around max pixel
• Standard deviation of pixel values	• Statistics for full image
• Median pixel value	• Statistics for row containing max pixel
• Difference between max and median values	• Statistics for column containing max pixel
• Difference between mean and median values	• Values for pixels in 5x5 region around max
	• Sorted values of pixels in 5x5 region around max

Figure 4.8: Contents of the small and large feature sets.

As expected, detection performance was best when the source was at the smallest standoff distance. At 30 meter standoff and 10 m/s driving speed an OOB error rate of 1% was obtained using the small feature set. This error rate was reduced to 0.65% when using the larger feature set. At 40 meter standoff the error rates were 7% and 4% for the small and large feature sets respectively. This pattern continues as the standoff distance increases, clearly demonstrating that including more information in the alarm signal can improve detection performance. A graph showing OOB performance as a

function of random forest tree size is shown in Figure 4.9. In order to determine how far the system could go without a false positive, a test set of 3992 background images was run through the classifier. Out of this set, 2 false alarms were registered. This suggests that the system would be able to drive around 399 kilometers between false positives after training on the 30 meter standoff, 10 m/s data.

4.6.6 Background subtraction in frequency domain

As the background distribution has a different overall shape than a point source distribution in coded aperture images, an interesting question is whether background subtraction is possible. To test this, the average 2D FFT of a small set of background-only images was calculated. In the same fashion, an average 2D FFT of source-only images at 30 meter standoff was also generated. Then, an unoptimized filter was created which emphasized the frequency components strongest in the source-only FFT. This filter was created using the average 2D FFTs from background and source only and manipulated by hand to give passable results by eye. The filter creation process can be seen in Figure 4.10, and an example of filtered vs unfiltered coded aperture images can be seen in Figure 4.11.

To test the effect of the filtering process, a random forest classifier was trained on each standoff distance using the large feature set as extracted from the filtered coded aperture images. A very significant increase in detection accuracy was observed as a result of the filtering process. The OOB error for 30 meter standoff at 10 m/s was reduced from 0.65% to 0.2%. At 40 meter standoff, the error was reduced from 4% to 1%. Using the filtered "30-10" settings, the same 3992 background images were tested with zero false positives. This suggests that the system could cover over 798 kilometers without false alarms while expecting a 0.2% error rate in detection.

The general accuracy of the random forest classifiers can be seen in Figure 4.12. A zoomed in version showing only the 30 and 40 meter standoff cases can be seen in Figure 4.13. It can be seen from these figures that the relative benefit of the filtering

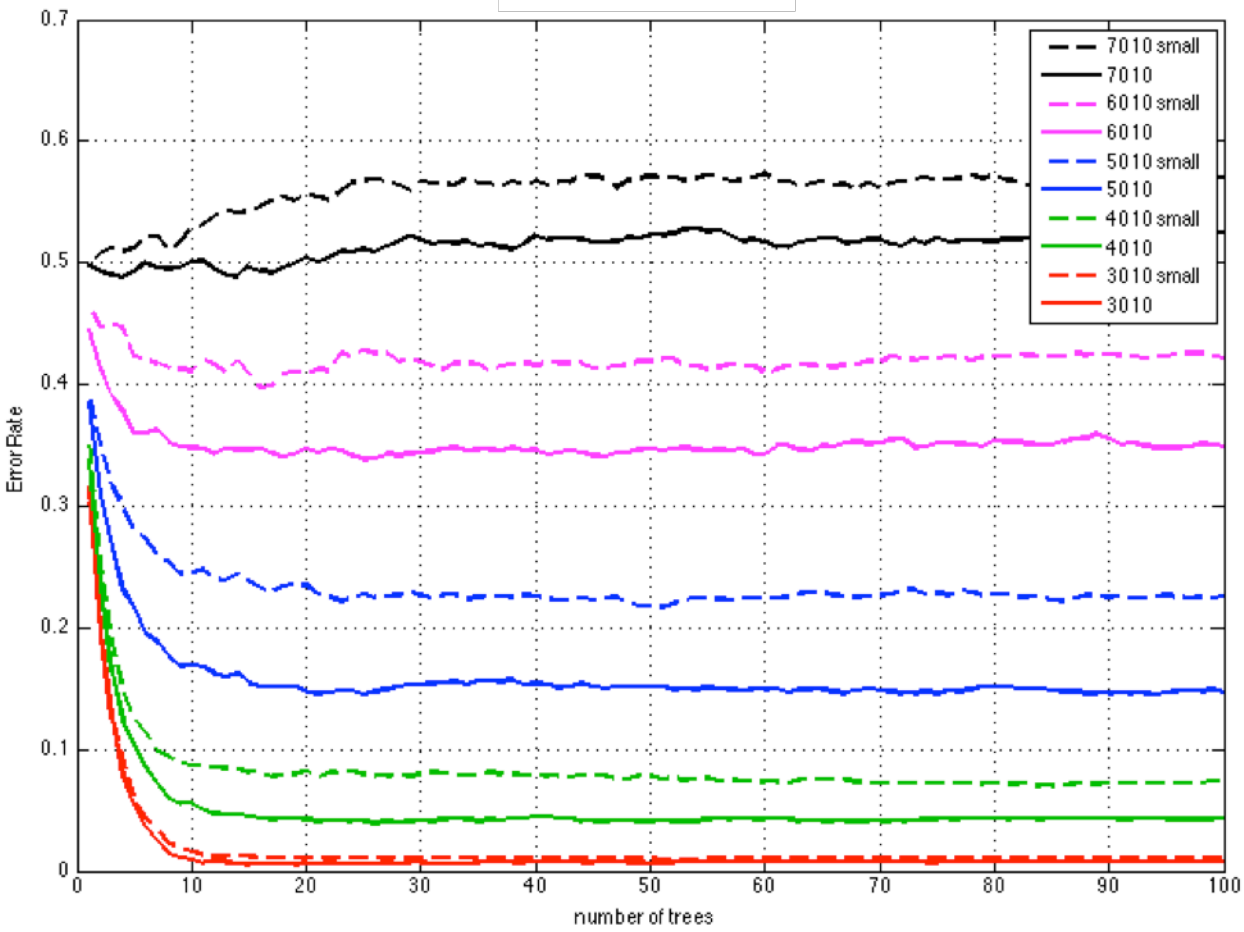


Figure 4.9: OOB error rates vs. forest size for random forest classifiers trained at different standoff distances. The dashed lines correspond to the small feature set. The solid lines correspond to the large feature set. In the legend, 3010 refers to 30 meter standoff, 10 m/s driving speed, and so on.

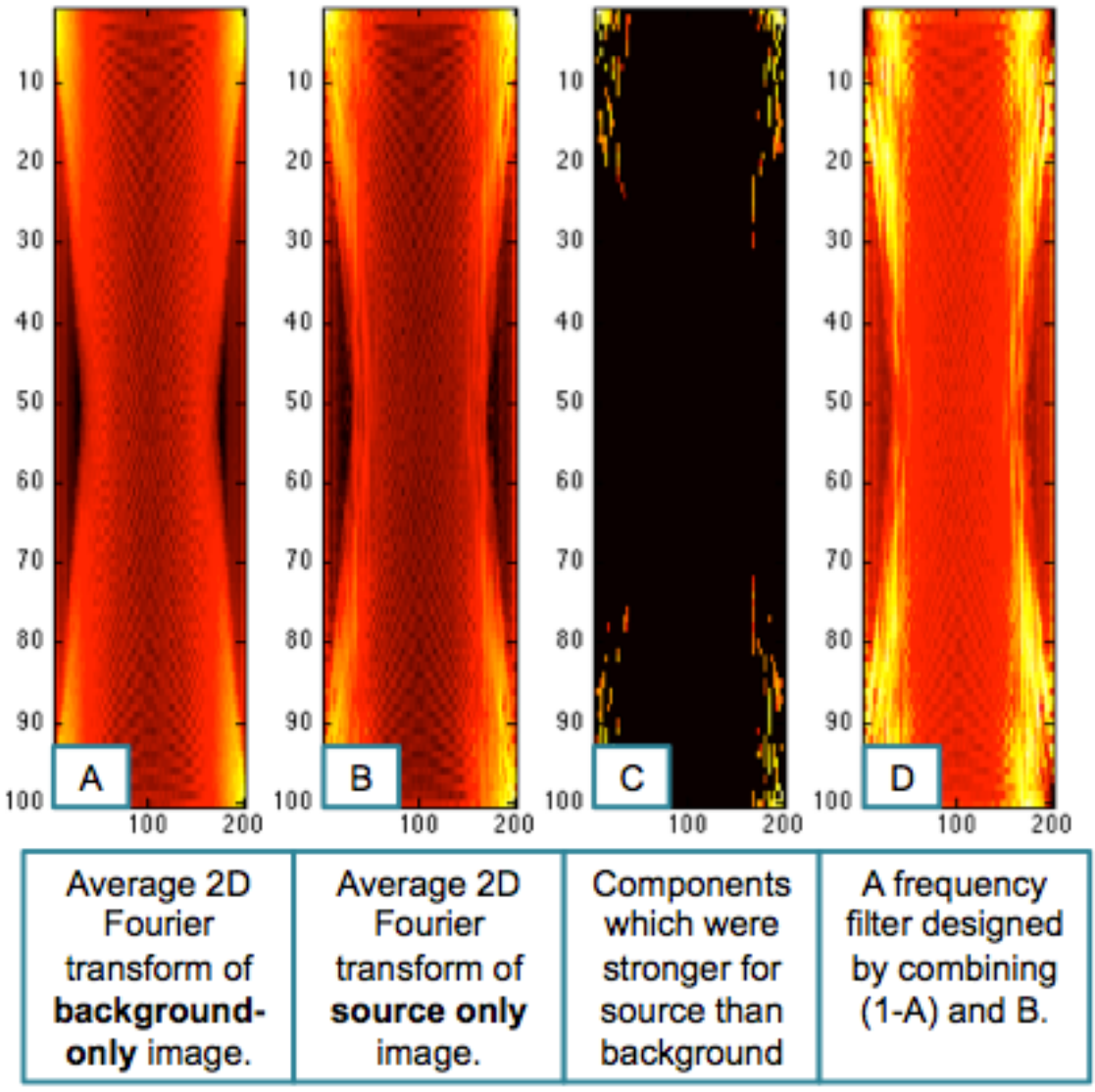


Figure 4.10: Design of background subtraction filter

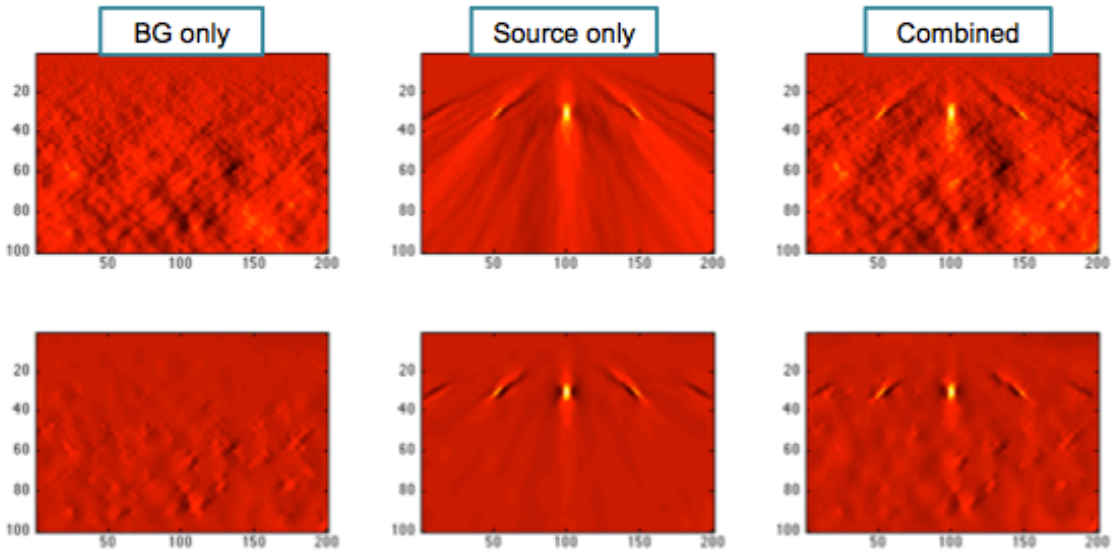


Figure 4.11: Effect of background subtraction filter. Top row shows unfiltered images. Bottom row shows filtered images. Columns from left to right show background-only, source-only, and combined images.

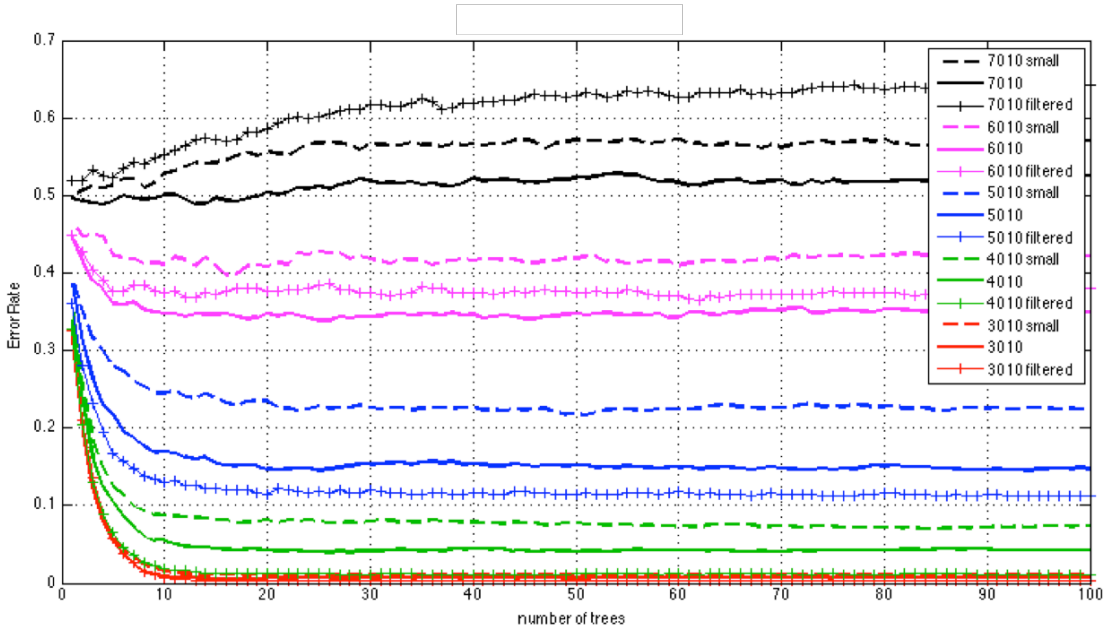


Figure 4.12: Effect of background subtraction. OOB error rates vs. forest size for random forest classifiers trained at different standoff distances. The dashed lines correspond to the small feature set. The solid lines correspond to the large feature set. The cross-hatched lines correspond to large feature set extracted from filtered images.

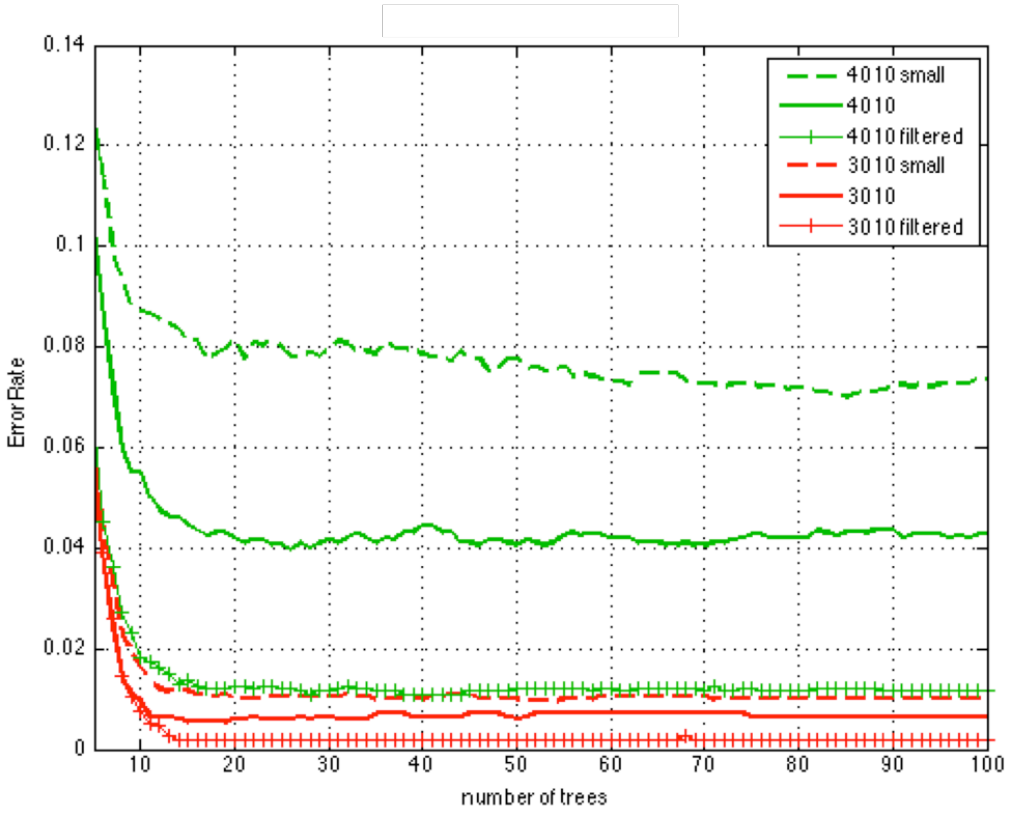


Figure 4.13: Effect of background subtraction. OOB error rates vs. forest size for random forest classifiers trained at 30 and 40 meter standoff distances. The dashed lines correspond to the small feature set. The solid lines correspond to the large feature set. The cross-hatched lines correspond to large feature set extracted from filtered images.

process decreases as standoff distance increases. Eventually, the filter leads to worse performance. This is likely due to the fact that the filter was designed using the 30 meter standoff source images. Another explanation could be that the source strength is so weak at the larger distances that throwing out any signal through the filtering process is detrimental.

4.6.7 PCA background subtraction

Some early background subtraction results with a hand designed frequency filter were shown in Section 4.6.6. Because this filter gave us a large performance boost, we may wonder whether we can do better using machine learning. Background subtraction for surveillance videos is a field where a considerable amount of research has been employed. Using sparse dictionaries or restricted Boltzmann machines, researchers have shown the ability to segment interesting objects out of images even when the background is non-stationary. This is accomplished by learning a background model and then generating a best guess for the background when presented with a new image. This generated background image is subtracted from the original, leaving only the foreground behind.

Further research will be required to understand whether background subtraction techniques designed for natural scenes are applicable to coded aperture imagery. In general, coded aperture intensity images have much less structure than a natural scene. Additionally, there is little consistency between a patch in one background image and the same patch in another background image. The main relationship is that the period of horizontal variations increases with reconstruction depth. Because of this, we may find texture based methods more useful.

It seems likely, however, that more structure can be introduced by moving from intensity-only to hyperspectral coded aperture images. This is because we could constrain the generated background to remain approximately the same "color", effectively preventing us from subtracting source components. Working with the

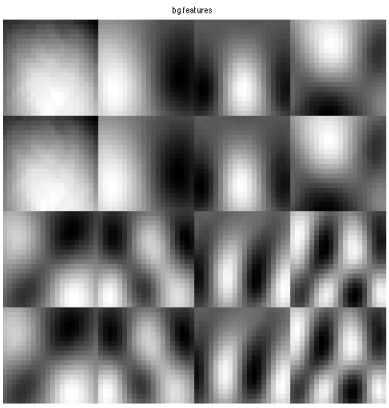


Figure 4.14: The first few PCA components from background-only images

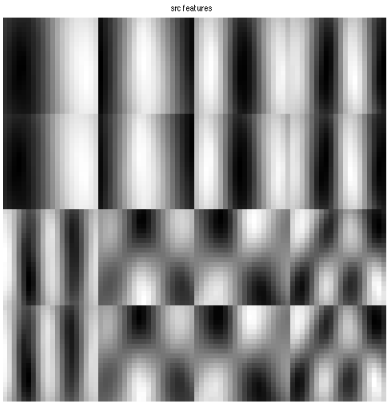


Figure 4.15: The first few PCA components from source-only images.

leftover image would be more effective than evaluating the spectrum of a single pixel because we can take spatial correlations into account. This can help reduce the false alarm rate from spurious pixels.

4.6.8 Background subtraction

PCA was used to learn components from the Large Area Imager source injection images. The images were split into tiles and the principal components were calculated. The first few principal components learned from background-only and source-only images for a 40×20 pixel tile size are shown in figures 4.14 and 4.15.

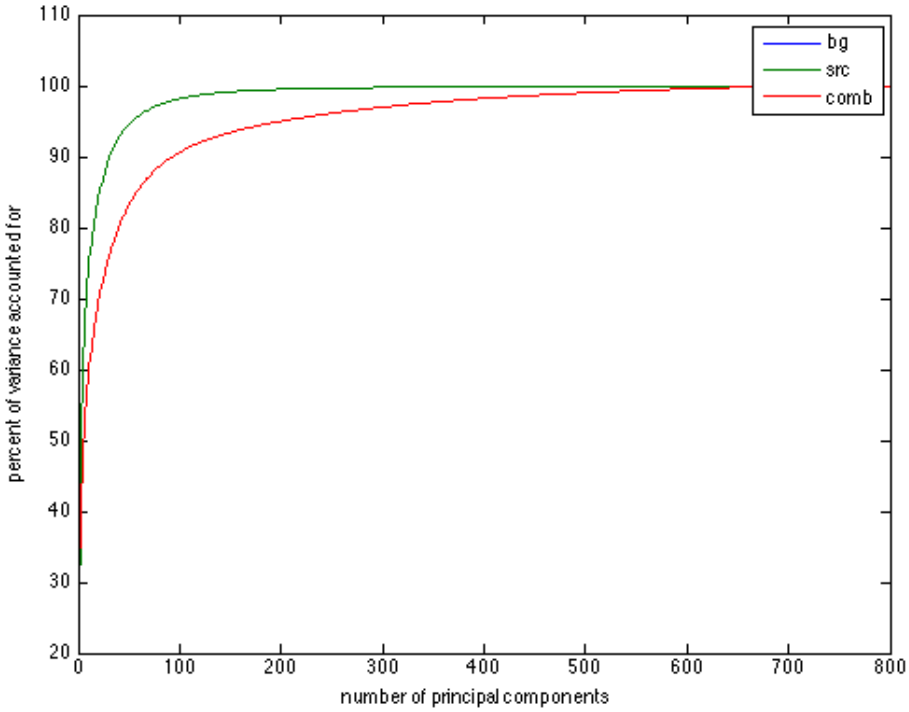


Figure 4.16: Percent of variation explained vs number of principal components. Results are shown for background-only (blue), source-only (red), and source-injected (green) 40×20 pixel tiles. The background-only line follows the source-injected line so closely that they are indistinguishable on this plot. Keep in mind that the principal components being enumerated are different for each tile set.

The percentage of the variance explained by the principal components is shown in figure 4.16. In this figure, the combined (injected) curve follows the shape of the background curve. This suggests that the sources shape has much less variance than the background. The relatively large number of components required to reach a high reconstruction rate in the background images is indicative of low correlation between pixels within a tile. This is not due to the fact that there is no structure at all, but comes from the fact that the tiling does not capture any regularity. This suggests that the pixel intensity may not be the best space in which to capture patterns in the background data.

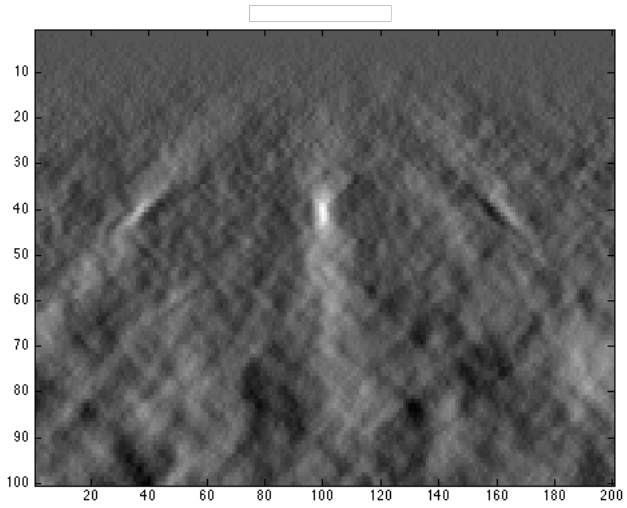


Figure 4.17: Source injected image before subtraction.

Background was subtracted by reconstructing a source injected image using only a few of the principal components from background only. The reconstructed image was then subtracted from the original image. This was done in the hope that only shapes common to the background would be well reconstructed by these first few principal components, leaving the source largely intact. Using the subtracted image as a multiplicative mask was also tried. The results of this subtraction are shown in figures 4.17 and 4.18 which show a source injected image and the image after masking, respectively.

Figure 4.19 shows a figure of merit (FOM) comparison using the estimated signal to noise ratio distributions calculated from the original, subtracted, and masked images. The figure of merit was calculated as:

$$FOM = \frac{\mu_2 - \mu_1}{\sigma_1 + \sigma_2}. \tag{4.25}$$

It can be seen that background subtraction is capable of improving the signal to noise ratio. Background subtraction with this method produces accuracies as good or slightly better than the hand designed filter described in Section 4.6.6. Masking

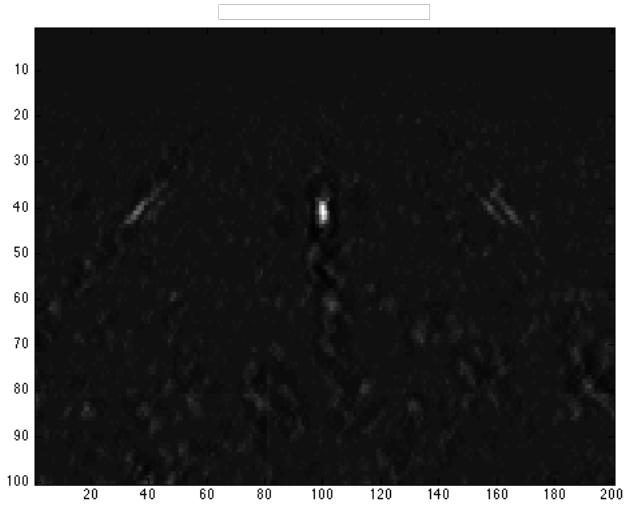


Figure 4.18: Source injected image after masking

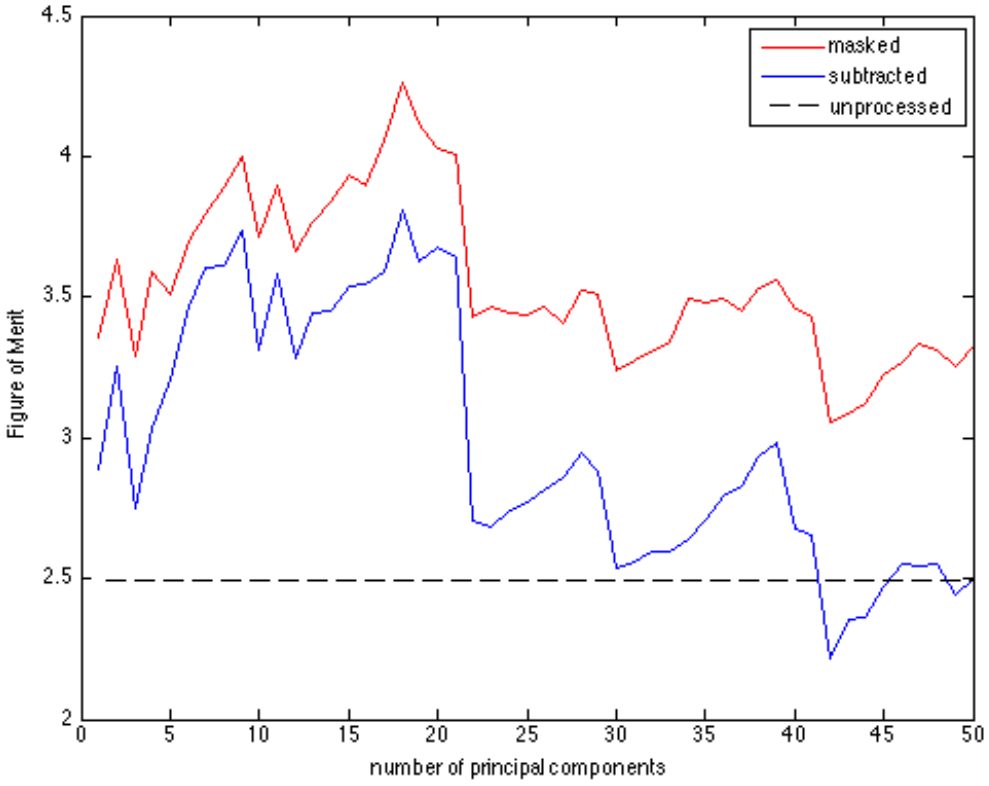


Figure 4.19: Figure of merit scores vs the number of principal components.

seems to improve the baseline figure of merit. This is likely due to an increase in contrast resulting from the multiplication. This benefit diminishes at larger standoff distances where the source does not show high contrast to the background.

The results of PCA background subtraction can likely be improved by switching to a Fourier or wavelet basis. The wave-like patterns are not spatially anchored within a background image. Thus random tiling will tend to wash out any relationships if the phase of these patterns is not allowed to vary.

4.6.9 Conclusions

A fairly realistic coded aperture performance study has been initiated using simulated sources injected into real measured background. These source injection images were used to train a random forest classifier whose output was used as an alarm signal.

It is clear that incorporating more features yields significant performance benefits. For the 30 and 40 meter standoff cases, the alarm signal error rate was cut approximately in half by moving from a small feature set with 8 entries to a larger feature set with 79 entries. This demonstrates that supervised classification algorithms are able to leverage the information encoded in the features despite the fact that they are highly correlated. The results of the 1D imaging study (Section 4.5) indicate that the choice of classification algorithm does not have an extreme effect on overall performance. Given these facts, feature extraction and discovery appears to be the most promising research area for low hanging fruit.

A second important conclusion is that background subtraction is possible in coded aperture images. Using a frequency filter designed to suppress components dominated by background, classification error for the 30 and 40 meter standoff cases was further reduced by a factor of three. Similarly, a generative PCA background subtraction method was able to improve a figure-of-merit for detection by approximately 175%. Results of this nature clearly demonstrate that the background, a distributed source,

is supported by a different set of spatial features than the point sources we seek to detect. We now have evidence that subtraction of these background features is not only possible, but that it can lead to significant improvements in detection capability.

Chapter 5

DDLI Proof of Concept System

5.1 Introduction

As mentioned in Chapter 2, neutron sensing was a major topic of interest for the sponsors of this work. Specifically, this meant understanding the potential benefits of incorporating neutron sensing into a mobile nuclear threat detection system. The sponsors acknowledged that gamma imaging technologies such as coded aperture had allowed for increased stand-off detection distances, directionality, and localization in previously constructed systems. The question was now, "Can transformational gains in detection capability be achieved via combined neutron and gamma sensing with similar technology?"

This part of the dissertation describes the design and build of a prototype coded aperture imaging system which incorporates both neutron and gamma detection capability. The discussion will begin with a review of related threat detection systems in Section 5.2. In Section 5.3, a series of results obtained early in the project is presented. With this context, a problem statement for the current work is outlined in Section 5.4. The build of the imaging system is described in Section 5.5. A series of mobile measurements conducted with the system are detailed in Section 5.6. Finally, results from the system are presented in Section 5.7.

5.2 Literature Review

5.2.1 Gamma Ray Imaging

By 2010, several gamma ray imaging technologies including coded aperture and hybrid scatter systems had been demonstrated in mobile platforms. These systems, designed to search for potential nuclear threat sources, leveraged imaging techniques to overcome fluctuations in the natural gamma ray background.

5.2.1.1 Large Area Imager

The Large Area Imager (LAI) was developed in 2003 by Klaus Ziock and others at Lawrence Livermore National Laboratory. It was a one-dimensional gamma ray coded aperture design originally based on NaI(Tl). In the stationary design, 57 NaI(Tl) detectors sat in a 3 × 19 array behind a linotype metal (85% Pb), base-19, uniformly-redundant array coded aperture mask. With a focal length of 1 meter and detectors of size 10 cm × 10 cm × 10 cm, the LAI was designed to provide 10 m position resolution at 100 m standoff distance. [1]

By 2007, the LAI had undergone significant changes. The system had been transitioned into a towable trailer to allow for mobile measurements. All NaI(Tl) had been replaced by 4.2 cm × 4.2 cm × 40 cm bars of CsI(Na) arranged in two 27-element detector arrays. Additionally, the mask design was now dual-sided, with one mask and one anti-mask section per side to allow for simultaneous mask/anti-mask imaging on both sides. This upgraded mobile design can be seen in Figure 5.1. [2]

In order to understand the performance of the imager, injectable sources were created from measured data [2]. This process involved measuring a high-activity Cs-137 source at specific distances and constant slow speeds. Using this data, it was possible to create a background-subtracted model of the source response at a given distance. Because coded aperture imaging is a linear process, the modeled source response could be realistically injected into any background dataset from the imager.

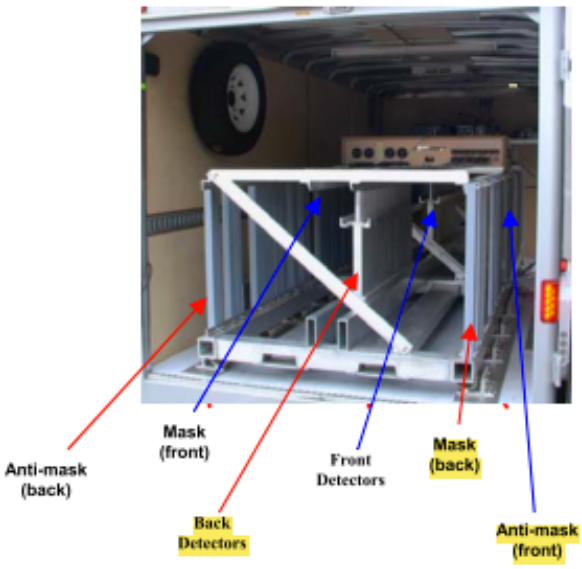


Figure 5.1: The 2007 Large Area Imager mounted in its trailer [2]. © 2007 IEEE.

This provided a way to simulate source measurements under varying background conditions given only a background measurement. A comparison between a measured and injected source can be seen in Figure 5.2.

Using the injectable source model, a source injection study was performed. In this study, sources were injected at different activities and distances into background data measured at over 2500 locations in the Livermore, California area. Alarms were issued by examining a 3×5 pixel location around the injection location and comparing the maximum SNR value for the Cs-137 peak to a set threshold. To understand the false-alarm rate, the same process was applied to background-only images at over 4000 locations.

The results of the source injection study are shown in Figure 5.3. The Large Area Imager was able to demonstrate very good detection performance at realistic driving speeds for Cs-137 at 50 meter standoff when the source activity was over 1.5 mCi. Similar performance was attainable for a weaker sources when the imaging distance was closer (40 m). These result showed that the imager could detect sources at greater distances and speeds than what had been demonstrated using counting methods.

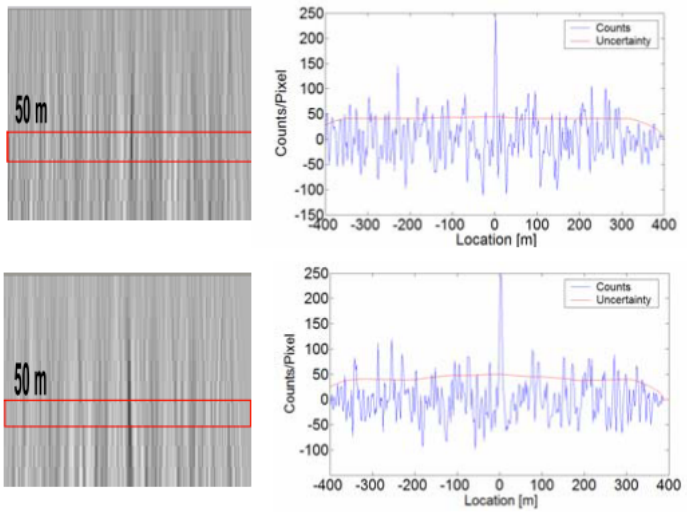


Figure 5.2: Top: An image formed passing a 1 mCi source at 50 m range. Bottom: An image formed when a 1 mCi model-source is injected into background data taken at the location of the top image. [2] © 2007 IEEE.

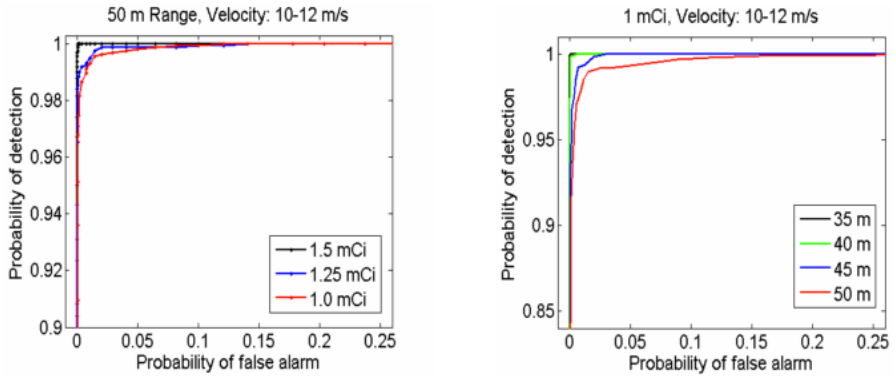


Figure 5.3: Left: ROC curves showing the performance of the LAI on injected data at 50 m standoff as source intensity varies. Right: ROC curves showing how distance affects performance when source activity is fixed at 1 mCi. [2] © 2007 IEEE.

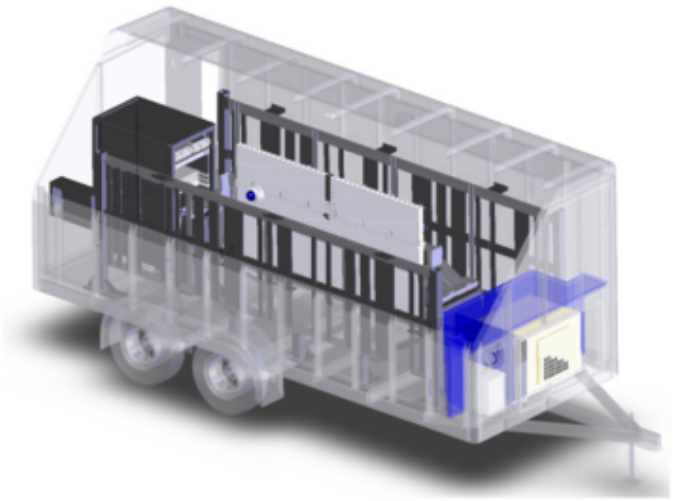


Figure 5.4: A model of the SORDS-3D system mounted in its trailer. [3]*

5.2.1.2 SORDS-3D

After the Large Area Imager measurements were complete, the readout technology was transferred to the Science Applications International Corporation (SAIC). SAIC then began designing a new system based on the Large Area Imager called SORDS-3D which was completed and tested by 2011. Like the Large Area Imager, SORDS-3D was dual-sided, with a rank-31 URA coded aperture mask on one side and the corresponding antimask on the other. Instead of two rank 27 element detector arrays, SORDS-3D had a single 37 element array of 5 cm $\times$ 5 cm $\times$ 50 cm, double-ended readout, CsI(Na) bars. The double-ended readout allowed for a vertical position to be established for an interaction using the light ratio of the two PMTs. Using this vertical position information in combination with a horizontal "shadow-gap" cut through the center of the mask, SORDS-3D is also able to estimate the elevation of a detected source. Unmodulated counts from this shadow gap reconstruct to a dark horizontal band in the image whose position is related to source elevation. A model of the SORDS-3D system in its trailer is presented in Figure 5.4. [3]

*Reprinted from Nuclear Instruments and Methods in Physics Research Section A, Volume 652, Penny et al. [3], "A dual sided coded aperture radiation detection system", pages 578-581, Copyright 2011, with permission from Elsevier

There were several algorithmic upgrades present in SORDS-3D. One major feature was an improvement in the way the image reconstruction was handled. Instead of simply projecting the mask pattern back into the world, the system response function was used to form the image. A ray tracing program was employed to determine the response function (photon detection probability) of the detector as a function of source position. This response function was then corrected to be zero sum by subtracting a weighted average no-mask and no-gaps response. These two responses are the detection probabilities if the imager had no coded aperture mask at all or a solid coded aperture mask with no gaps, respectively. This process can be seen in Figure 5.5.

The response function formulation employed by SORDS-3D was designed to suppress background contributions and artifacts compared to the unweighted projection. Additionally, a point by point adjustment to the response was used to allow the imager to continue suppressing artifacts even when making turns. This gave SORDS-3D the ability to project into a geospatial frame and overlay over normal maps. The approximately 6% energy resolution of CsI(Na) at 662 keV was also exploited. Energy spectra, built for each pixel in the projected image, were passed to the RadioNuclide Analysis Kit (RNAK) any time the pixel's activity exceeded a 4σ threshold. RNAK uses template matching to solve for the best matching spectral constituents using non-negative least means squares. [3]

Although exact performance values could not be published due to United States export control restrictions, SORDS-3D was able to identify sources in the milliCurie range at distances up to 130 meters and has been tested at speeds up to 95 km/h [3]. It is likely that the algorithmic improvements of SORDS-3D are responsible for its remarkable performance, as the underlying detector technology is relatively unchanged from the Large Area Imager.

[†]Reprinted from Nuclear Instruments and Methods in Physics Research Section A, Volume 652, Penny et al. [3], "A dual sided coded aperture radiation detection system", pages 578-581, Copyright 2011, with permission from Elsevier

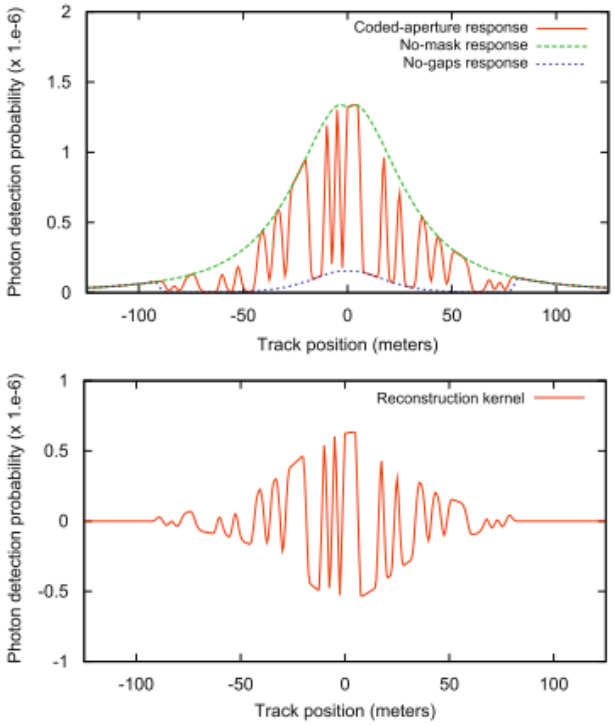


Figure 5.5: Top: The 35 meter standoff response for a single SORDS-3D detector as calculated by the ray tracing program. Bottom: The zero sum kernel formed from the three response functions in the top image. [3][†]

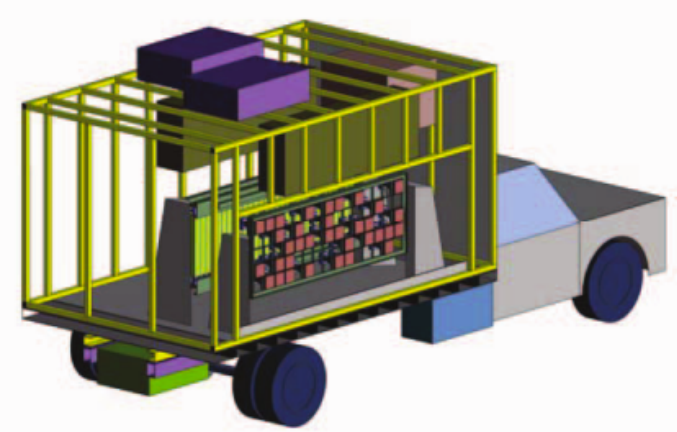


Figure 5.6: A model of the Trimodal Imager. [4] © 2009 IEEE.

5.2.1.3 Trimodal Imager

The Raytheon-SORDS Trimodal Imager (TMI) was developed in 2009 by Raytheon, Los Alamos National Lab, and other partners. In contrast to 1-D coded aperture systems like the Large Area Imager and SORDS-3D, the Trimodal Imager was designed to do 2-D coded aperture imaging. Instead of lead, the Trimodal Imager employed 5 inch × 5 inch × 2 inch NaI(Tl) detectors as an active coded aperture mask. This active mask sits in front of an array of 2.5 inch × 3 inch × 24 inch NaI(Tl) bars which form the coded aperture detector plane [4]. The inclusion of an active mask plane allowed the Trimodal Imager to simultaneously perform both Compton scatter and coded aperture imaging. The imager also extracted non-imaging data from a technique called shadowing detection for early alert indication [42]. A model of the Trimodal Imager can be seen in Figure 5.6.

The Trimodal imager used an image fusion technique to combine the coded aperture and Compton scatter images it produces. The imager built energy spectra for the source pixels, provided depth information, and could overlay a detection location over the view provided by a side mounted camera.

Studies conducted using validated simulations [4] of the Trimodal imager and randomly varied simulated background showed the increased performance attainable

from fusing the coded aperture and Compton imaging modalities [42]. As expected, the performance of Compton imaging increased with increasing gamma ray energy while coded aperture imaging dominated at lower energies where the mask was most opaque. There was a notable increase in performance gained from the image fusion process. Interestingly, this improvement was present even in the Cs-137 study where the accuracy of Compton scatter imaging alone was comparable to a coin toss [42].

5.2.1.4 MISTI

MISTI stands for Mobile Imaging and Spectroscopic Threat Identification. The MISTI system was developed at the Naval Reasearch Laboratory in 2009. It was designed to combine the high energy resolution available from high purity germanium (HPGe) detectors with coded aperture imaging based on more affordable NaI(Tl) detectors. The system design goal was the ability to detect 1 mCi of Cs-137 at a range of 100 meters in 20 seconds. [5]

The MISTI coded aperture detector plane was a 10×10 array of 10 cm × 10 cm × 5 cm NaI(Tl) detectors. The coded aperture mask was a 50% coverage 12×18 pseudo random pattern made of lead tiles. In addition to the coded aperture array, 28 p-type, 110%-efficiency ORTEC HPGe detectors were mounted in the MISTI truck. These HPGe detectors provided spectroscopy with resolutions around 2 keV FWHM at 1333 keV. A model of the MISTI system is shown in Figure 5.7. [5]

In a departure from the other gamma ray threat detection systems presented here, MISTI's alarm algorithm is completely based on spectroscopy. Imaging only occurs after an alarm has been issued and is used to localize the source. To detect a source, spectra are histogrammed in 1, 2, 4, 8, 12 and 20s sliding accumulation windows. These histograms are updated every second and sent to peak analysis software packages. The MISTI looks at peaks in two ways. The first method is a Region of Interest (ROI) method which gates around known peaks using a library. The second is a Second Difference Peak Search performed by Genie2K. Detection

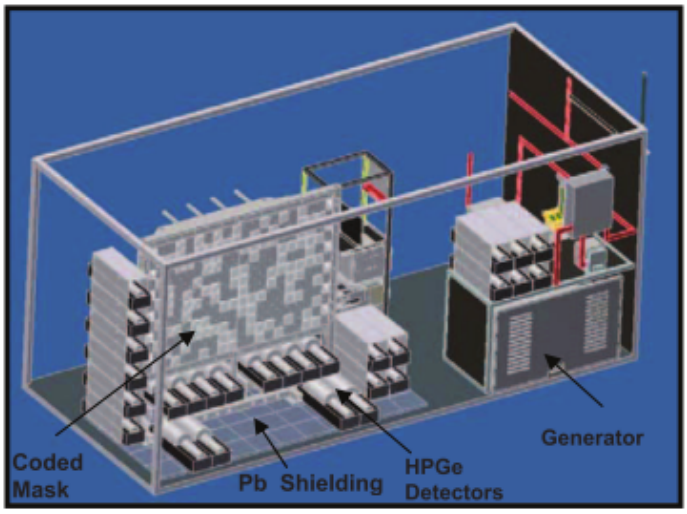


Figure 5.7: A CAD model of the MISTI system. [5] © 2009 IEEE.

thresholds are set such that background can generate only one false alarm every eight hours. Once a peak is detected an identified, a coded aperture image can be made using the energy lines of the detected isotope. [5]

Methods of improving performance in the face of background were a major focus of study with the MISTI system. A cross country background study was conducted by NRL around 2011 [43]. This study looked at the variation in background lines and intensity as the imager travelled, considering the effect this might have on false alarms.

In 2014, after the transfer of MISTI to the University of California, Berkeley and Lawrence Berkeley National Lab, detailed background surveys in the San Fransisco Bay Area were taken with the MISTI system. These background surveys were used to perform a source injection study aimed at understanding the effect of background fluctuation on spectroscopic and coded aperture detection. It was found that any methods which reduced background-induced variation improved detection performance [6]. Spectral background estimation techniques were demonstrated to improve detection performance significantly. In the study, an optimized background subtraction estimate using neighbor windows outperformed all other results. It was

also demonstrated that coded aperture imaging always outperformed spectroscopy with a simple energy window and no background subtraction [6]. Results from the study may be found in Figures 5.8 and 5.9.

5.2.2 Neutron Imaging

5.2.2.1 Neutron Scatter Camera

The first prototype of the Neutron Scatter Camera was completed in 2006 by Sandia National Laboratory. Designed to image fission spectrum neutrons, the Scatter Camera worked on a principal similar to Compton scatter imaging.

The prototype imager was composed of two planes. In the front plane were four 2.5 cm diameter, 5 cm thick EJ301 liquid scintillators. The back plane contained, four more EJ301 detectors with a diameter of 13 cm and a thickness of 5 cm. The prototype was able to successfully image an AmBe source over a wide field of view and reconstruct the source spectrum by summing energy depositions in both planes [7]. An example image can be seen in Figure 5.10

By 2007, the detector size and number had been significantly increased and several algorithmic improvements had been introduced. These include the ability to discriminate gamma rays from neutrons using a 2-D decision boundary in PSD-TOF space and the introduction of Compton Scatter imaging for gamma rays [8]. An example of the PSD-TOF cut is shown in Figure 5.11. By 2011, the number of detectors had been increased to 24 and MLEM reconstruction was being used in place of back-projection [9]. A photo of the device as it existed in 2011 is shown in Figure 5.12.

5.2.2.2 ORNL Fast Neutron Imager

The ORNL Fast Neutron Imager was developed by 2009 at Oak Ridge National Lab. Designed to do 2D coded aperture for fission-spectrum neutrons, the device consisted of 25 fast plastic scintillators in a 5 $\times$ 5 configuration behind a high-density

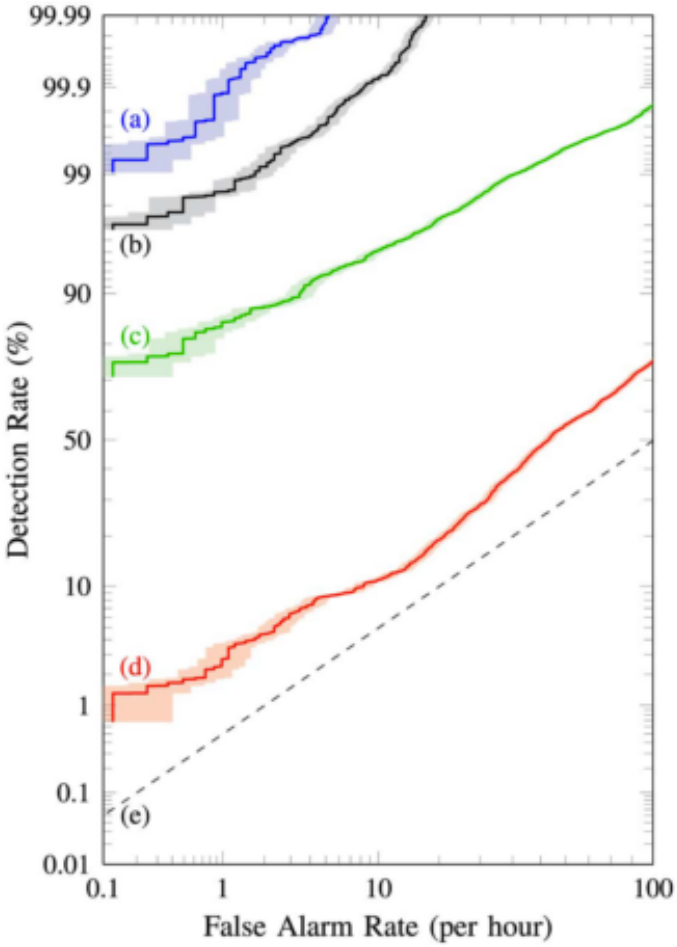


Figure 5.8: Source injection ROC curves for detecting 1 mCi of Cs-137 with MISTI at 100 m standoff and 25 mph. These results are for spectroscopic and windowing algorithms: (a) Statistical limit, (b) Trained spectroscopic estimate, (c) Linear spectroscopic estimate (d) No background estimate, (e) No sensitivity. Shaded regions indicate 68% confidence intervals. [6] © 2014 IEEE.

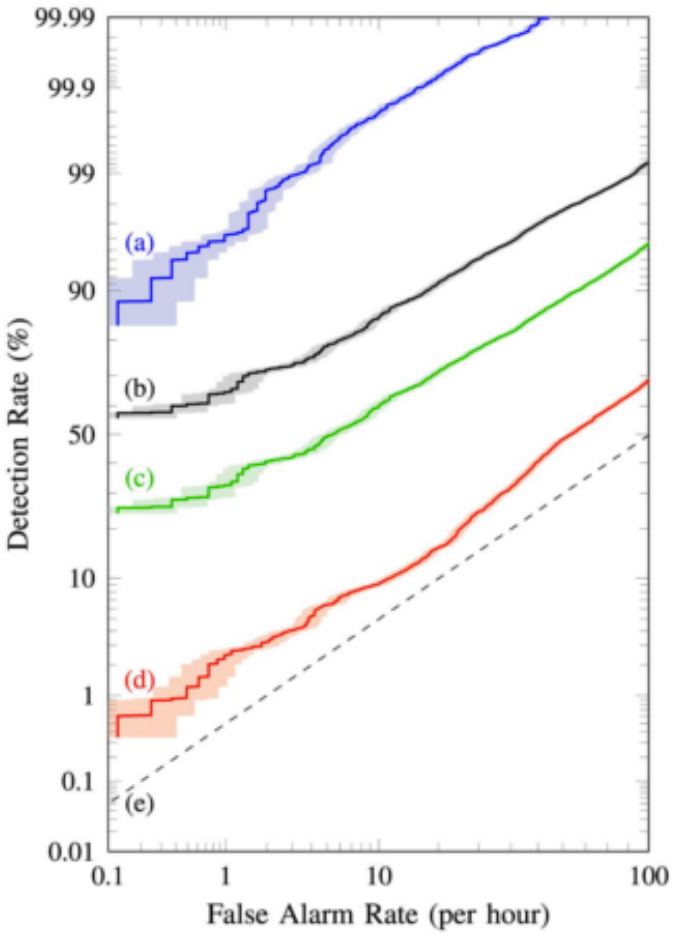


Figure 5.9: Source injection ROC curves for detecting 1 mCi of Cs-137 with MISTI at 100 m standoff and 25 mph. These results are for coded aperture algorithms: (a) Statistical limit, (b) Ideal image, (c) Decoded image, (d) No imaging, (e) No sensitivity. Shaded regions indicate 68% confidence intervals. [6] © 2014 IEEE.

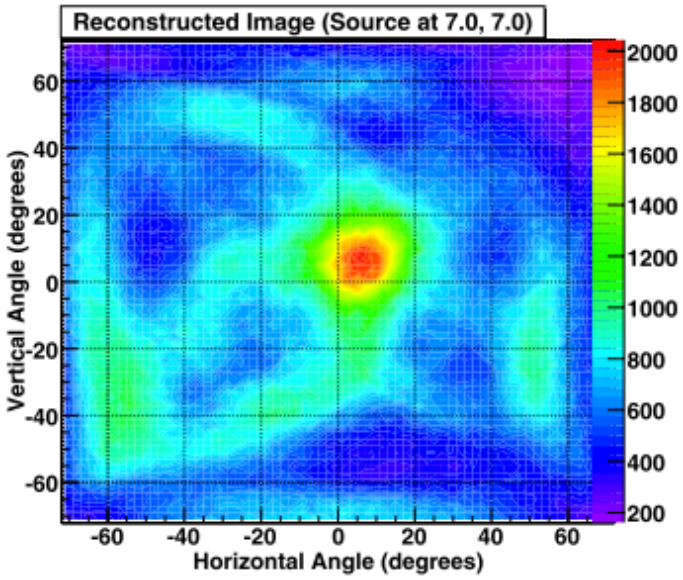


Figure 5.10: An image of an AmBe source produced by the 2006 Neutron Scatter Camera prototype. [7] © 2006 IEEE.

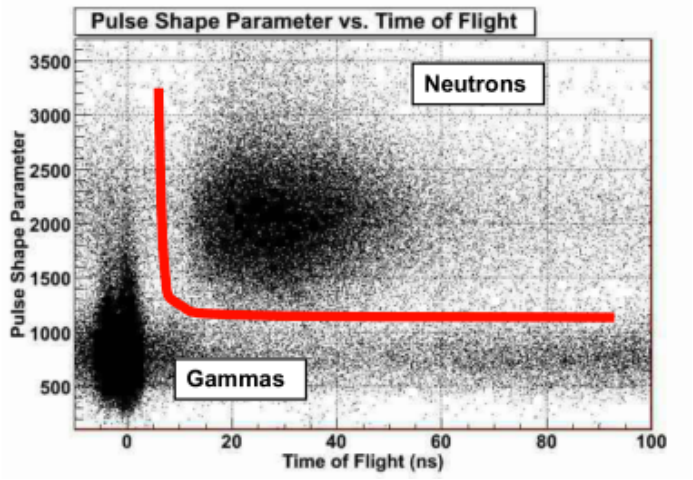


Figure 5.11: Neutron-gamma discrimination in the Neutron Scatter Camera is a function of both pulse shape discrimination and time of flight. [8] © 2007 IEEE.

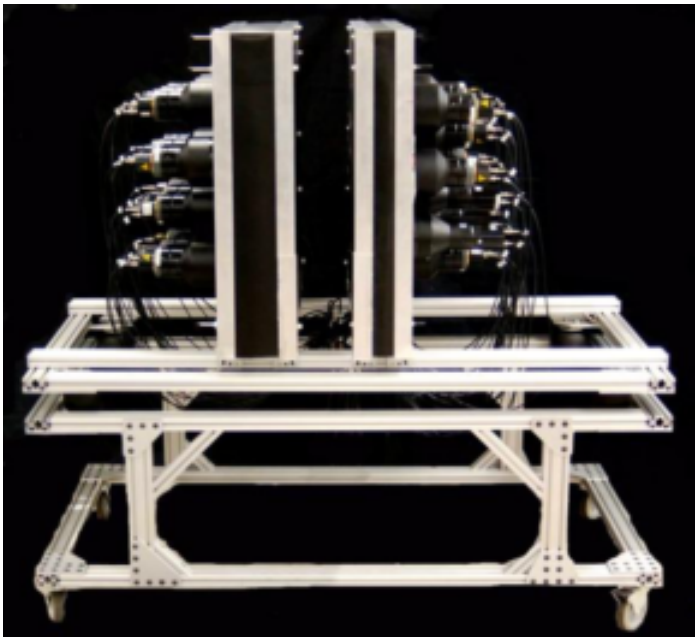


Figure 5.12: A side view of the Neutron Scatter Camera in 2011. [9]

polyethylene, rank-19, 2-D MURA coded aperture mask. The detectors themselves are pixelated. Each one contains a 10 × 10 array of plastic scintillator pixels backed by a light guide and viewed by four photomultiplier tubes read out using Anger logic. The acquisition system for the system was adapted from a Siemens Positron Emission Tomography (PET) system with 0.312 ns timing resolution [10]. A photo of the imager can be seen in Figure 5.13.

The ORNL Fast Neutron Imager has been tested with plutonium mixed oxide (MOX) fuel pins at the Idaho National Laboratory ZPPR facility. Measurements were performed with the system in both passive and active modes. For the passive measurements, MOX fuel pins were placed in two configurations, one a "soup can" filled completely with pellets, and the other a "clamshell" containing nine pins arranged in the form of an "I" [10]. Images generated in passive mode of each object are shown in Figure 5.14.

For the active measurements, highly enriched uranium (HEU) plates containing 8.785 kg of U-235 were placed in a clamshell. U-235 does not emit neutrons spontaneously, so a D-T neutron generator was used to induce fission in the material.

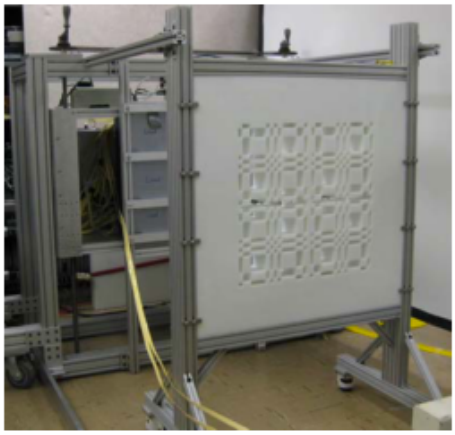


Figure 5.13: A photo of the ORNL Fast Neutron Imager prominently featuring the 2-D coded aperture mask. (Image reproduced with permission from "Passive and Active Fast-Neutron Imaging in Support of AFCI Safeguards Campaign" by Paul Hausladen and Matthew Blackston.) [10]

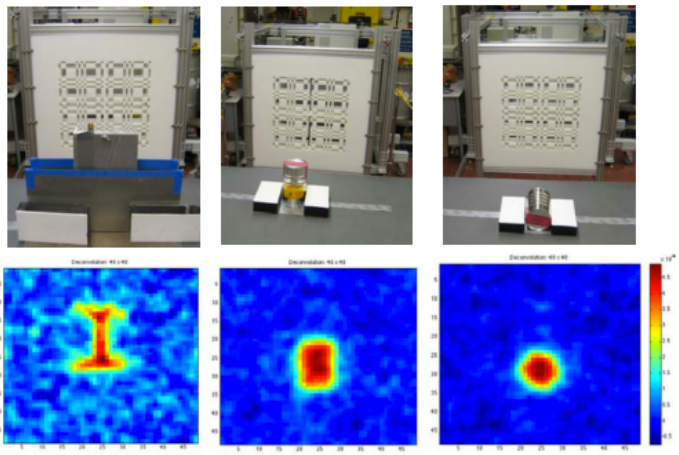


Figure 5.14: Photographs of the imaging configuration along with coded-aperture neutron-gamma images from the ORNL Fast Neutron Imager. (Left) The clamshell with "I" configuration. (Middle) The soup can standing vertically. (Right) The soup can laying horizontally. (Image reproduced with permission from "Passive and Active Fast-Neutron Imaging in Support of AFCI Safeguards Campaign" by Paul Hausladen and Matthew Blackston.) [10]

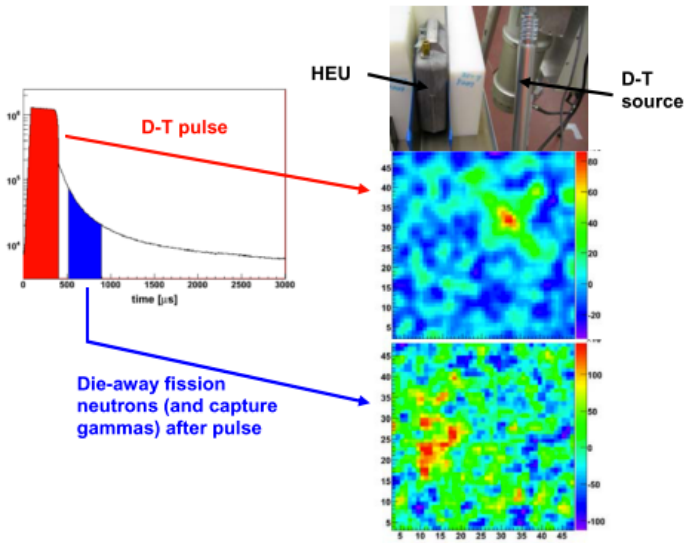


Figure 5.15: Active imaging with the ORNL Fast Neutron Imager. The left graph shows the counts in the imager with respect to D-T generator pulse time. On the right, three images are shown. The material configuration (top), an image of the D-T pulse (middle) and an image of the material from die-away neutrons and capture gammas (bottom). The bottom image is a difference between measurements with and without the HEU to remove background. (Image reproduced with permission from "Passive and Active Fast-Neutron Imaging in Support of AFCI Safeguards Campaign" by Paul Hausladen and Matthew Blackston.) [10]

Moderating material was placed around the material on three sides to bring the 14 MeV D-T neutrons to thermal energies where fission is more probable. Timing was used to ignore the neutrons from the original D-T pulse and look at fission neutrons and capture gammas in a die-away image. Figure 5.15 shows the results of this measurement. This was the first demonstration of die-away imaging with fast neutrons [10].

Because of its good spatial resolution and efficiency, the ORNL Fast Neutron Imager design could be applied to situations where special nuclear material needs to be characterized in place. This includes situations like holdup measurements, treaty verification, and storage monitoring [10].

5.3 Early Work

The DDLI project, which the work presented here concerns, was a 5 year project. In order to understand the design choices in the system build and measurements discussed later, it is important to describe results achieved earlier in the project. This section will cover a range of topics including the specific topics of interest of the project; the design iterations leading to the proof of concept design and their motivations; experiments to understand neutron background in urban environments; and studies of neutron imaging vs. neutron counting. In addition to the information presented here, a machine learning study related to the DDLI system and the development of a supporting Geant4 software toolkit are described in Chapter 4 of this document.

5.3.1 System Design Iterations

This section will briefly describe the design iterations of the DDLI system. These designs will be referenced throughout the following discussion of early project work.

In the original proposal, the detector and mask designs were a heterogeneous mixture of detector materials. During this phase, some consideration was given to thermal neutron sensing in addition to gamma rays and fast neutrons. To facilitate this, boron loaded liquid scintillators and CLYC were incorporated. Additionally, the mask plane contained cadmium to modulate thermal neutrons. Liquid scintillator is also capable of detecting fast neutrons and gamma rays. CLYC also provided the ability to detect gammas. HPGe facilitated high gamma ray energy resolution. A diagram of this design can be seen in Figure 5.16.

During Phase I of the project, which took place during 2010 and 2011, two new system designs were investigated. The first design was one-sided, featuring a lead and liquid scintillator active mask and a dual-layer detector plane. The second was a double sided design featuring a single, stacked detector plane and cadmium in the mask. Models of these designs are shown in Figures 5.17 and 5.18.

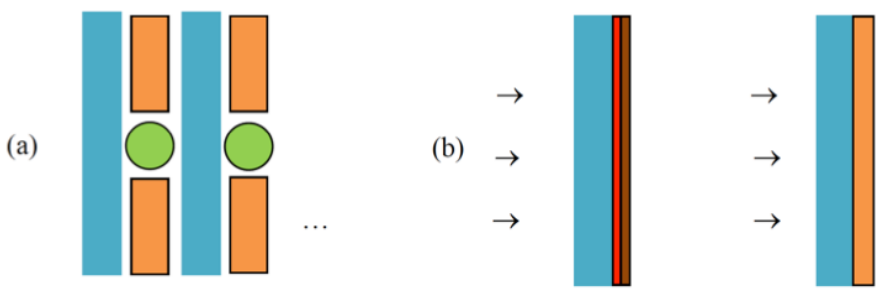


Figure 5.16: (a) Front view of detector plane. The proposed system envisioned boron-loaded liquid scintillator (blue), CLYC (orange), and HPGe (green). (b) Mask concepts 1 (left) and 2 (right). Arrows indicate direction of incident radiation. Mask concept 1 includes a thin Cd layer (red) and Pb alloy layer (brown).

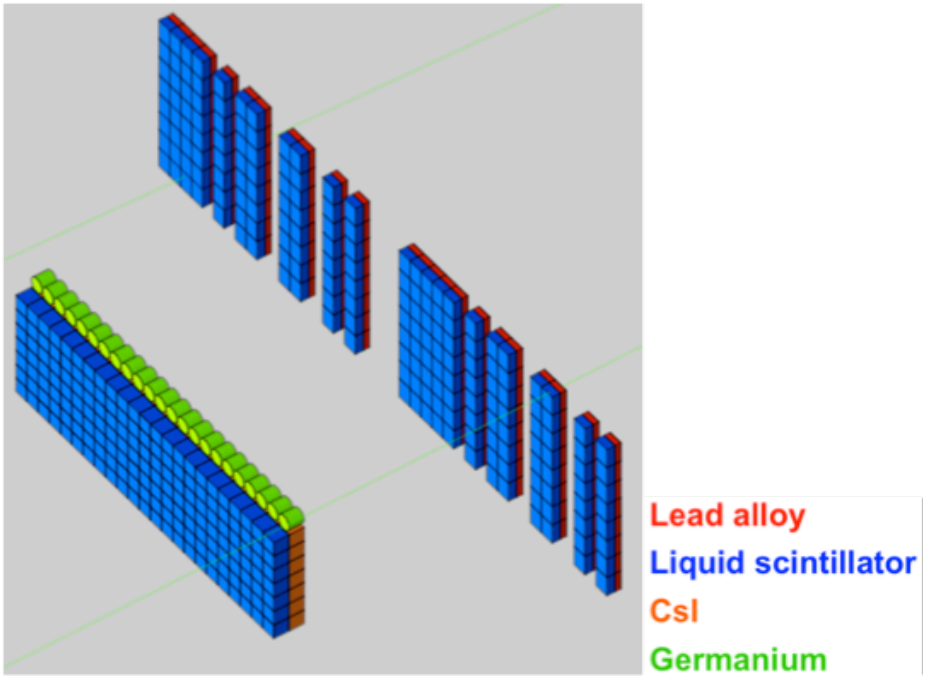


Figure 5.17: Single-sided concept from Phase I with layered detector plane.

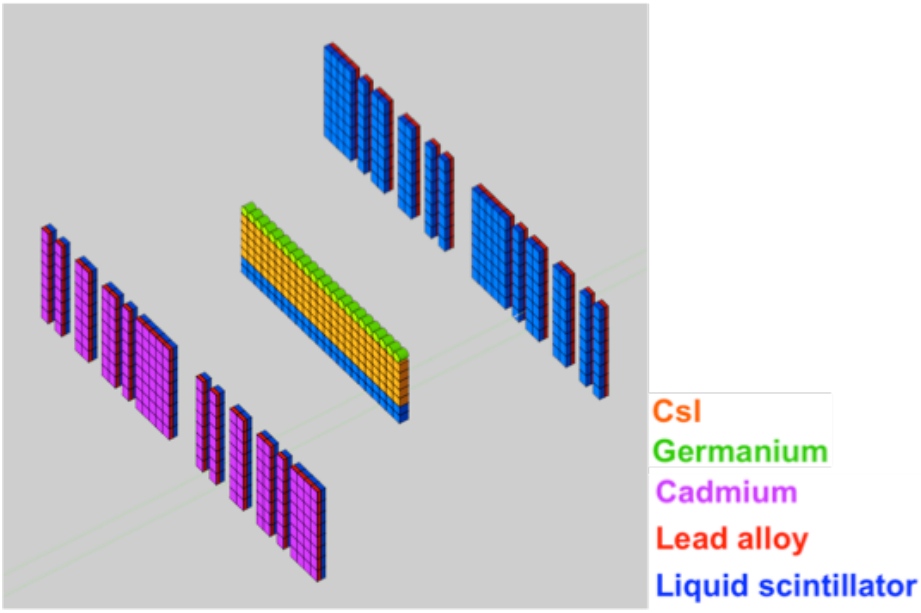


Figure 5.18: Dual sided concept with a single detector plane layer from Phase I.

In Phases II and III (2011-2012), an optimized full system model was used for simulation. This model's parameters were chosen based on simulated optimization studies for various detection metrics. The optimized design featured a dual-sided, rank 29 active neutron, passive gamma mask with 0.4 mm of cadmium and a focal length of 119.6 cm. The detector plane contained 4 rows of 8 cm × 8 cm × 8 cm CsI detectors; 3 rows of 8 cm × 8 cm × 12 cm organic scintillators; and 1 row of 8 cm × 8 cm HPGe detectors. A model of this design can be seen in Figure 5.19.

As mentioned in Section 2.1, a proof of concept (PoC) design was expected to be built and demonstrated as part of this project. To realize a PoC, a scaled down version of the DDLI was designed. This PoC version can be seen in Figure 5.20. The DDLI PoC features a single-sided, active neutron, passive gamma mask with the capability to perform coded aperture imaging for both gamma rays and fast neutrons. Due to cost constraints, a single row of plastic scintillators is present in the mask, with the remaining portion of the neutron modulating mask filled in by high density polyethylene. Mechanically cooled HPGe assemblies are mounted such that the high energy resolution detectors are unmodulated by the mask. The system is mounted in

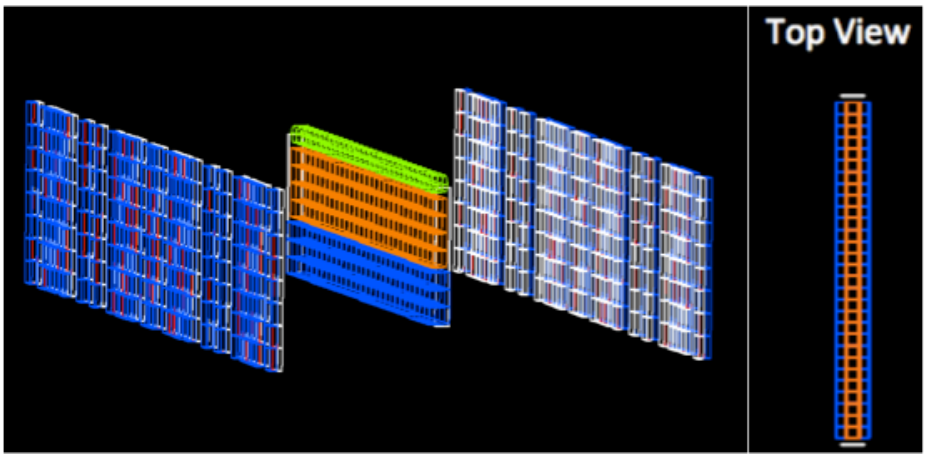


Figure 5.19: The optimized design used in simulations during Phases II and III.

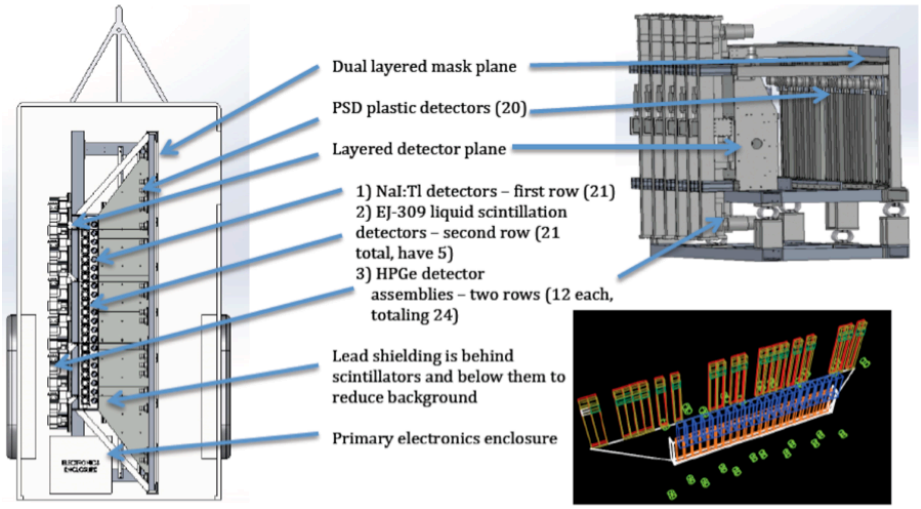


Figure 5.20: The original design for the DDLI proof of concept build.

a towable trailer. This trailer is the same one that was used to house the Large Area Imager (see section 5.2.1.1). The front of the trailer has space for a large generator. System electronics are housed in a rear enclosure. A fully instrumented model of this PoC system without HPGe can be seen in Figure 5.21. Finally, the operating version of the DDLI PoC can be seen in Figure 5.22. In this version, which was actually constructed, the HPGe detectors were removed.

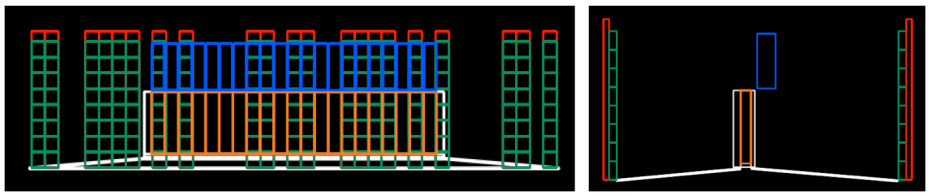


Figure 5.21: The scaled up PoC design featuring two, fully-instrumented, active coded aperture masks.

Figure 5.22: The version of the PoC which was actually built and tested. This is a partially instrumented version without HPGe.

5.3.2 Neutron Counting vs. Imaging

Early in the project, it was assumed that neutron imaging would provide benefits over neutron counting in the same way that gamma ray imaging does. For this reason, most analysis in Phases I-III focused on neutron coded aperture imaging as the primary neutron detection modality. In Phase III of the project, a neutron counting study was conducted to mirror the earlier neutron imaging investigations. When coupled with a neutron background measurement campaign in downtown Knoxville, TN, these results indicated that neutron counting could outperform neutron coded aperture imaging in the optimized DDLI system. This is especially true in cases where the neutron background is fairly well behaved.

In Phase II, a simulated study focused on coded aperture imaging was conducted. To understand the detection performance of neutron coded aperture in the optimized DDLI system, 100 passes of the imager past a 1 significant-quantity (SQ) weapons grade plutonium (WGPu) source were simulated in Geant4. These were injected into 100 passes including simulated background. This yielded 100 source plus background

(S+BG) and background (BG) images, whose distributions were used to produce ROC curves for detection performance. This process was repeated at several standoff distances. As an approximation to a PSD threshold, neutrons depositing less than 1 MeV in the scintillators were rejected.

To capture the qualities of the fast neutron background, which originates from the interaction of GeV cosmic rays in the atmosphere, background neutrons were modeled according to published results on cosmic ray induced neutron background at sea level according to MCNPX [41]. This model included directionality, energy dependence, and neutron albedo from the ground with good agreement to measurement. Importantly, the intensity of the neutron background did not vary with time in these simulations.

Results from the simulated neutron coded aperture study are shown in Figure 5.23. Fairly poor neutron detection performance was indicated by simulations when using neutron coded aperture, even at a very low speed and standoff distance (30 meters). Compare this result to Figure 5.24, which show ROC curves for neutron counting with and without an active mask at 30 meter standoff using the same simulation workspace. These results, generated in Phase III, show that neutron counting allows for significantly improved detection performance even at 10 times the speed.

One major assumption in the treatment described above was that the intensity of neutron background did not vary during the simulated measurement. According to a measurement campaign conducted in downtown Knoxville, TN [11], the neutron background does show fluctuations. Although these background fluctuations are not as severe as those seen in the gamma background, they can still have a significant effect. Figure 5.25 shows background suppression as observed during several static dwells along a downtown street. This chart shows the fractional intensity of the neutron background relative to measurements taken outside of downtown away from large buildings. This effect is strongly correlated to the angle of open sky, which

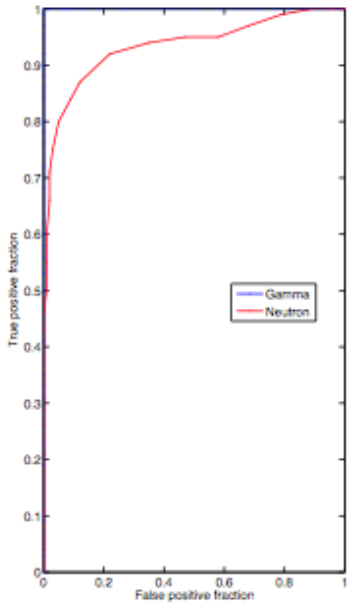


Figure 5.23: A ROC curve for simulated imaging with 1 SQ of WGPu at 30m even at 1 m/s. Note the poor performance of neutron coded aperture even at this slow speed.

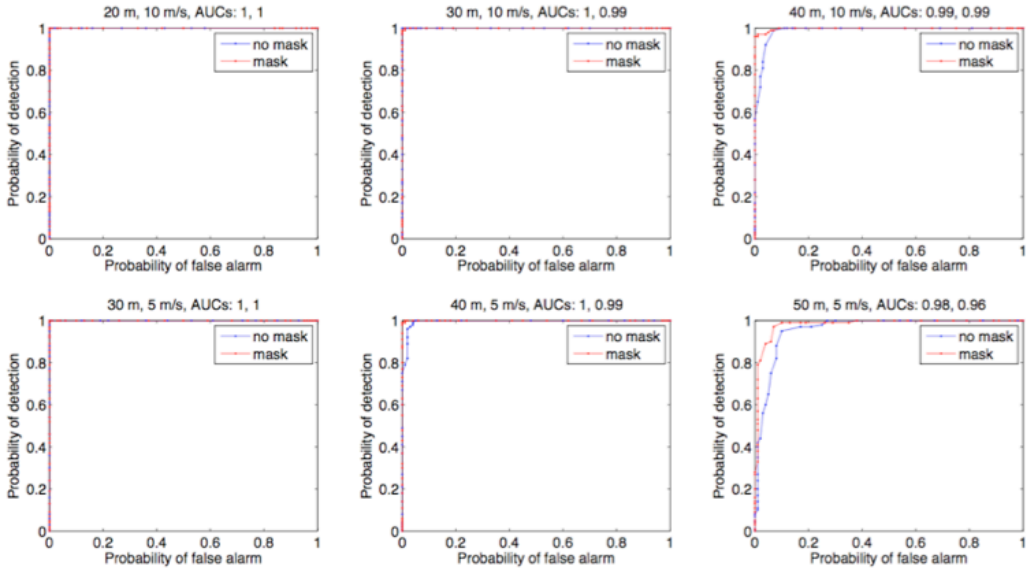


Figure 5.24: ROC curves for simulated neutron counting in EJ-309 liquid scintillator for 1 SQ WGPu with and without an active mask. Note that performance at 30 m standoff is significantly better than obtained with imaging (Figure 5.23) even at 10 times the speed.

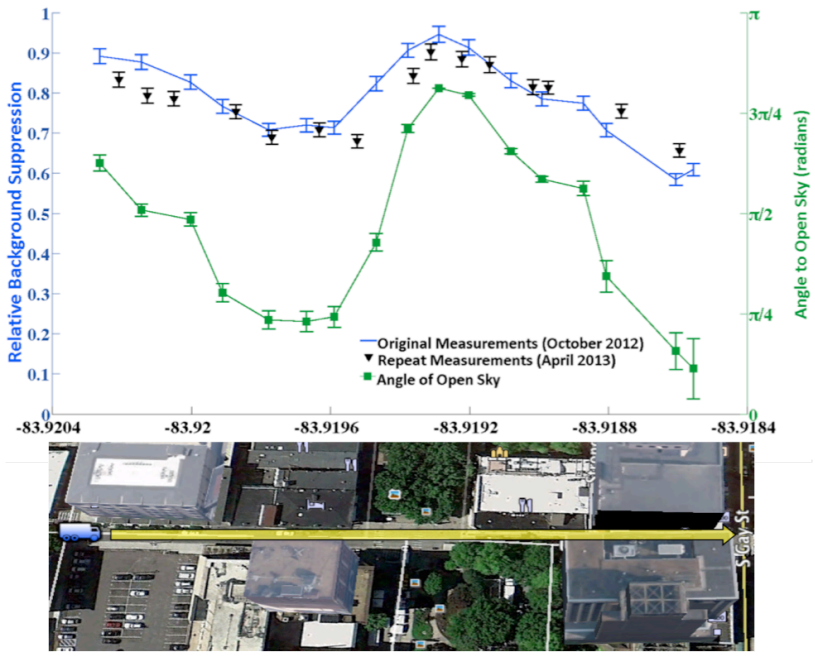


Figure 5.25: Top: Background suppression in downtown Knoxville, TN and its correlation with occlusion of sky by surrounding buildings. Bottom: A satellite view of the street approximately lined up with the above graph. [11][‡]

indicates how much of the largely downward directed neutron background is blocked by structures. A mean neutron suppression of approximately 20% was observed in the downtown Knoxville area.

To understand the effect of a variable background on neutron counting, random dips with a mean suppession of 20% were added to the mean neutron count rate to generate a time dependent background signal. This is shown in Figure 5.26, with and without statistical variation. These large, building-like dips had a significant impact on the neutron counting detection performance. This result is visualized in Figure 5.27, where detection rates are reduced to below 50% for reasonable false alarm rates.

Considering the strong correlation shown in Figure 5.25, it may be possible to correct for this suppression given some sort of open sky metric, e.g. from video feed. However, this problem is unsolved at the moment. Correction techniques applicable

[‡]Reprinted from Nuclear Instruments and Methods in Physics Research Section A, Volume 773, Iyengar et al. [11], "Systematic measurement of fast neutron background fluctuations in an urban area using a mobile detection system", pages 27-32, Copyright 2015, with permission from Elsevier.

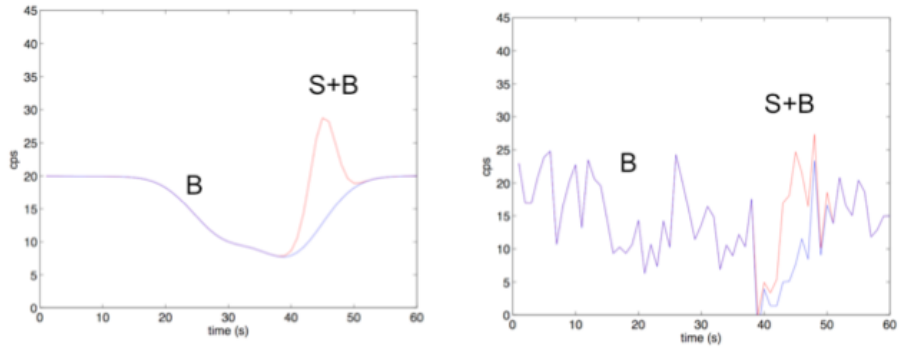


Figure 5.26: Left: A simulation showing a large neutron contribution from a source convolved with a 20% dip in the neutron background due to local suppression. Right: The same data with the addition of poisson counting fluctuations.

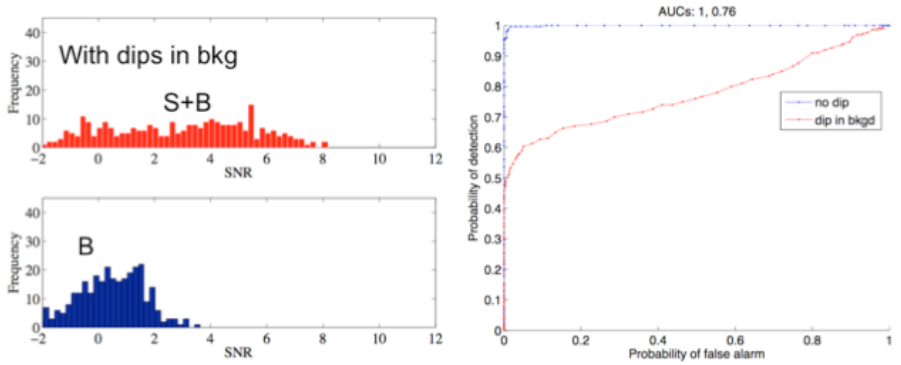


Figure 5.27: ROC curves (on right) for neutron counting detection with and without neutron background suppression.

to gamma ray counting, such as moving average windows, are likely to be ineffective for the neutron background due to the low count rates involved. Neutron background fluctuations like the ones that have been shown are only discernable over a time frame of minutes in a position. A moving imager, however, would only spend a few seconds in each location.

5.4 Problem Statement

This section will describe the overall tasks remaining during the final stages of this project and how they were addressed. This will constitute a statement of the problem motivating the experimental work of this dissertation. First, the project state – as it existed when experimental work began – will be contextualized by the goals described in Chapter 2. In an effort to help the reader understand the course of the PoC build, measurements, and results, the original plan for project completion will be detailed in Appendix B. Associated experiments are discussed in Appendix C.

5.4.1 Project State

Following the early work described in Section 5.3, it was time to develop a plan for the proof of concept (PoC) build and demonstration. Early simulation results had helped guide the development of an optimized DDLI system, and further studies clarified what was most important when considering a PoC design. Additionally, neutron background studies had informed us about the variability of the background and some metrics that might be used to correct it.

At that time, many of the goals outlined in Section 2.1 were complete within the simulation space. A mobile imager had been designed which was capable of simultaneous gamma-ray and neutron coded aperture imaging. Although no new materials were included in this design, their effectiveness and cost had been considered in the earlier simulation studies. The waveform digitizers present in the

new system would enable for gamma/neutron separation based on PSD within the organic detectors. Timing resolution was also good enough to correlate interactions for enable scatter imaging. Advanced algorithms had been investigated within the simulation workspace (see Chapter 4), including methods to suppress background clutter in coded aperture imagery.

What remained at this point in the project was a physical execution of this simulated work through a PoC build. This would involve the construction and housing of the DDLI PoC system, wiring detectors, configuring software and readout systems, managing electrical power, maintaining climate control, system configuration and more. Following the intensive build process, it would then be important to characterize the performance of the system in a way which aligned with the project's research goals. This would require detailed planning and execution of experiments and subsequent data analysis.

In large part, the PoC effort would focus on demonstrating the effectiveness of neutron sensing as a detection modality. Gamma ray systems had been extensively investigated in the threat detection community. As the utility of gamma ray imaging and spectroscopy was not in dispute, the most useful and novel course would be design measurements assessing the benefit added by the PoC's fast neutron detectors.

To summarize, the remaining tasks for the build portion of the DDLI project were as follows:

1. Construct the DDLI PoC system
2. Bring the DDLI PoC system online and configure it for operation
3. Design and execute a measurement campaign with a focus on neutron sensing
4. Analyze measured data and report on system performance

The completion of items (1) and (2) will be extensively detailed in Section 5.5. Measurements with the PoC system are described in Section 5.6. Finally, system performance results are presented in Section 5.7.

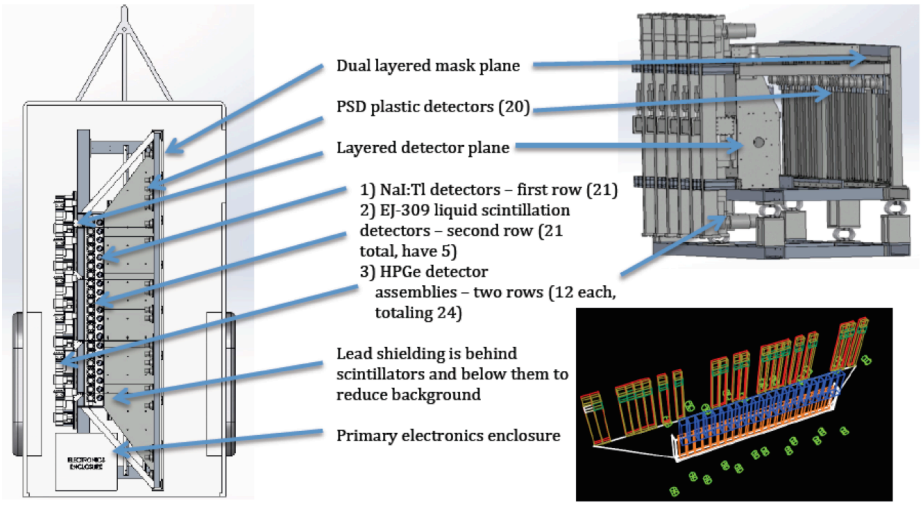


Figure 5.28: A SolidWorks model of the DDLI PoC system with all detectors mounted.

5.5 Build Process

5.5.1 Trailer and Frame

In May 2014, the Large Area Imager was shipped in its trailer from SAIC San Diego to ORNL. It had been determined that using the existing Large Area Imager trailer and generator would result in an estimated savings of $37k for the project. The PoC frame, which had been designed to fit in this trailer, had begun construction in February 2014 in advance of this shipment. This aluminum frame would house all of the system's detectors and coded aperture mask and took approximately 8 weeks to build.

The frame was designed to accomodate a full array of 21 NaI(Tl) bars, 21 dual-PMT liquid scintillators, 1 row of plastic scintillators, 24 mechanically cooled HPGe detectors, a lead coded aperture mask for gamma rays, a polyethylene coded aperture mask for neutrons, a lead ground shield, and an electronics enclosure. A SolidWorks model of the frame with all detectors mounted can be seen in Figure 5.28.

The DDLI PoC frame was mounted in the LAI trailer in June 2014 after removing the existing Large Area Imager system. 16 out of 24 HPGe detectors (8 of 12

Figure 5.29: A photo of the DDLI PoC frame and HPGe detectors mounted in the trailer.

assemblies) were also loaded into the trailer at this time. An photo of the mounted frame is shown in Figure 5.29 After mounting the HPGe detectors, the system tongue weight was over 1430 lbs, which exceeded the capacity of the existing LAI trailer hitch.

5.5.2 Detector Installation and Wiring

By September 2014, the detectors and electronics chasis with shock absorbers had been mounted into the trailer. A VME crate was installed in the electronics chasis along with an array of cards:

- 1 embedded Intel Core I7 VME Controller with 2 TB SSD removable storage
- 6 Struck SIS 3316 250Ms/s, 14-bit, 16-channel waveform digitizers
- 5 CAEN HV 6533 6-channel negative high voltage supplies

All organic scintillators were connected to high voltage power and the digitizers by September 2014. The five liquid scintillators used two channels of high voltage

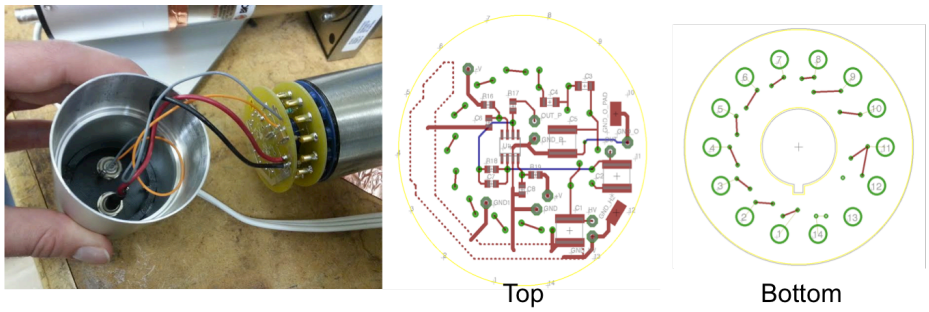


Figure 5.30: A photo of a complete NaI base on a detector (left) along with a schematic of the two circuit boards that make up the assembly (middle and right).

and one digitizer block each. The plastic scintillators, which contained an onboard potentiometer, shared one channel of negative high voltage between a pair of detectors.

The NaI detectors were equipped with PMTs, but had been obtained without bases. It was determined that the most cost effective solution would be to design and manufacture the bases for these units in-house. Using a design by Dr. Lorenzo Fabris, 25-two layer circuit boards were fabricated. Soldering for these boards was done by a contractor and a grad student within the group. The boards were mounted into a 3D-printed housing and wired to standard connectors. The bases were then attached to the detector pins and held in place with copper tape. A photo of the circuit board and housing can be seen in Figure 5.30 along with a board schematic. This process was completed by October 2014, at which time the NaI(Tl) detector bases were connected to a standalone CAEN positive high voltage supply and the VME digitizers.

5.5.3 Troubleshooting

5.5.3.1 HPGe Removal

In June 2014, after significant effort to bring the HPGe cooling electronics into compliance with ORNL electrical safety specifications, 8 of the 12, two-detector HPGe assemblies were mounted into the DDLI trailer. At this time, the HPGe coolers were

connected to trailer power and allowed to cool down for 24 hours. Once they were cool, bias and preamplifiers were connected to the detectors in order to perform an initial test with an oscilloscope and MCA.

Unfortunately, out of all the assemblies, only one detector from one pair was found to be in working order. As the systems have a temperature dependent bias interlock, it was concluded that the mechanical coolers were functioning properly. It appeared that the vacuum seals on the detectors had been compromised, rendering them useless. According to the manufacturer, Canberra, these assemblies were designed with a higher number of vacuum boundary seals than is typical. Due to high cost, Canberra and two other HPGe manufacturers were unwilling to service them.

Without any other avenues, the HPGe assemblies were offloaded from the trailer to save weight. One assembly was taken to UTK for an attempted fix with no allotted budget, but any success would have no bearing on this project. The DOE property transfer for the HPGe assemblies was also halted to avoid paying use tax on nonfunctioning equipment. A lead counterweight was added to the trailer to alleviate weight imbalance from removing the extremely heavy HPGe assemblies from the design.

5.5.3.2 Plastic Scintillator Bases

As mentioned previously, the plastic scintillators were equipped with a potentiometer. This potentiometer was included in the design to allow for a pair of plastic scintillators to share one channel of negative high voltage. In this scenario, gain matching would be carried out by adjusting the brightest detector in a pair down with the pot. The motivation for this scheme was the reduced cost in terms of high voltage channels.

After the installation of the organic scintillators, the first attempt at gain matching was undertaken. At this time, it became clear that the manufacturer, Eljen, had installed a potentiometer but had completely bypassed it on the board. As shipping the detectors back to Eljen could take a significant amount of time, it was decided that we would fix the plastic detectors ourselves.

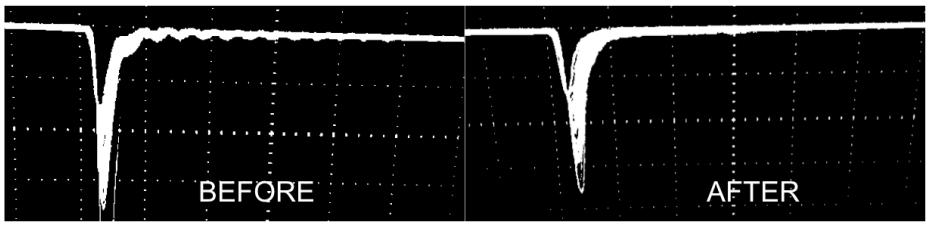


Figure 5.31: Waveforms from a plastic scintillator. After enabling the potentiometer, significant ringing artifacts appeared (left). The signal was cleaned up by terminating the dynode internally (right).

In October 2014, all 20 plastic scintillators were disconnected from their wiring and mounts and brought into the lab. As a simple fix, the potentiometer was disconnected from ground and wired to affect the high voltage bias. However, with the potentiometer enabled, the detectors exhibited significant ringing in the signal. The first attempted correction for the ringing problem was to add a 1 MF capacitor to each base between the potentiometer and ground. This seemed like it corrected the problem, so the fix was applied to all detectors. However, this fix was falsely attributed to the capacitor and the ringing remained.

It was later found that making a connection to the dynode eliminated the ringing. The false attribution stemmed from the fact that the dynode output was connected briefly by accident before testing the first detector. This connection was found to temporarily reduce anode ringing even after disconnecting from the dynode. Using this information, the ringing was finally eliminated by terminating the dynode output. A 50-ohm parallel connection was made between the dynode and ground, and the resulting clean signal can be seen in Figure 5.31.

5.5.3.3 Overheating and Under-heating

Attempts at running the full trailer system were met with almost immediate failure. During the first gain matching measurements, the trailer's data acquisition program terminated itself after only a few seconds. Initially suspecting a bug in the software, the source code was scoured to find the responsible line. Far from a bug, a safety

featured was discovered. As it turned out, the maximum safe temperature for operating the SIS3316 digitizer cards was 52° C. Reaching this threshold would result in an immediate halt command from the readout software. Surprisingly, this temperature was being exceeded.

The overheating condition was surprising for two reasons. The first was that the VME crate for the DDLI PoC was purchased specifically to avoid overheating conditions. Previous projects within the ORNL group had encountered overheating, so this was a known issue. As such, a crate which was advertised to have superior cooling capability was installed from the start. The second reason for surprise was that the DDLI PoC trailer was air conditioned. It was expected that the air conditioning unit could keep the trailer cool enough to avoid problems, especially without the mechanical HPGe coolers and electronics.

Upon inspecting the air conditioning system, it was found to be leaking water from the vents. The trailer was taken to an A/C specialist, who discovered that it had frozen up. The technician found the maximum cooling level which the trailer air conditioner could sustain and returned it. Unfortunately, even with the air conditioner functioning, data acquisition was still failing after a short time.

The only remaining option at this point was to upgrade our VME crate's forced convection system. A pastic plenum was designed and manufactured using a 3D printer. A section was cut away from the metal top panel of the crate and the plenum was installed. Two powerful, independently powered fans were attached to the plenum to suck air past the VME cards. These were then outfitted with 3D printed finger guards, as they are located in a high traffic area of the trailer. With the addition of these fans, data acquisition was finally able to continue. Card temperatures had been stabilized in around 40° C. A photo of the fans and plenum can be seen in Figure 5.32.

By this time, winter was fast approaching. We were now concerned with the opposite heating problem. If the temperature in the trailer dropped too low, we could have problems with optical coupling in the detectors. Even without these

Figure 5.32: The high power fans mounted to the 3D printed plenum.

problems, swings in temperature would cause gain fluctuations, which we would like to minimize. To keep the trailer warm, three ceramic panel heaters were affixed to the inside wall. These were connected to a thermostat which would would prevent them from heat the trailer too much.

Even with heating panels installed, however, the temperature in the trailer was reaching levels too low for comfort, especially at night. This was due to the complete lack of insulation in the trailer, which was only made of the wooden floor and the metal outer paneling. To improve the effectiveness of our heaters, insulating boards were installed along the walls and ceiling of the trailer between the support arches. Despite leaving the floor as it was, the insulation effort was extremely effective. Both heating and cooling were much improved by this addition.

What had not been improved was the project timeline. The first air conditioning issues were discovered in November 2014. The truck insulation addition was completed in February 2015. Diagnosing and fixing these temperature control issues had taken approximately four months of project time.

5.5.4 Configuration Software

Data acquisition in the DDLI PoC trailer is performed by Dr. Jason Newby's NGM software system. This software is capable of talking to the digitizer cards, writing

their buffers to binary files, and post processing those binary files into ROOT trees [44]. At the start of data acquisition, the digitizer cards must be configured to match the requirements of each detector channel and set various system options. This configuration is typically performed by executing configuration commands in a Python script before starting a measurement run.

Because the NGM software writes to registers in the VME cards to set options, most stages of the configuration required the user to input hex values corresponding to a set of on and off bits. Sometimes individual bits in a register represented boolean options. Sometimes ranges of registers were used to represent integer values for parameters. Often these parameters had machine units which were not intuitive. Overall, this process was highly error prone and could require frequent journeys into the NGM source code and digitizer documentation.

To make configuration easier and less prone to measurement-invalidating errors – which the author was guilty of frequently – a set of helper tools was created in Python. These helpers attempted to bring configuration into a more human readable realm.

The first tool was a Python class called DetectorMapping. DetectorMapping is an interface to an XML database of detector information. It provides a way to query for specific criteria, for example, finding all the liquid scintillators in the system. This tool was useful because it allowed for a significant amount of reusable information to be separated from the configuration script itself. Information from an XML database can be updated in one central location and called into any script to set up a measurement without the need to update many configuration files if something changes in the system. In our case, database entries contained information on the type, location, position, digitizer channel, high voltage channel, and optimized bias for each detector.

The second tool was a Python class called ConfigHelper. ConfigHelper used metaprogramming and properties to provide human readable and intuitive ways to set digitizer options. ConfigHelper is built of several different types of properties which abstract the NGM register manipulation functions. Unit conversion properties

```
# original configuration snippet
for islot in range(0,cards.GetEntries()):
  card = cards.GetValueO(islot)

  for iblock in range(0,4):
    card.firthresh_block[iblock] = 150
    card.firenable_block[iblock] = 1
    card.gate_window_length_block[iblock] = 500
    card.sample_length_block[iblock] = 500
    card.sample_start_block[iblock] = 0
    card.pretriggerdelay_block[iblock] = 250
    card.risetime_block[iblock] = 4

    for ichan in range(iblock*4,(iblock+1)*4):
        card.trigconf[ichan] = 0x5
        card.firenable[ichan] = 1
        card.fircfd[ichan] = 0x3
        card.risetime[ichan] = 4
        card.gaptime[ichan] = 0
        card.termination[ichan] = 1
```

```
# with DetectorMapping + ConfigHelper
for detector in mapping.match({"Type":"Liquid"}):
  slot = int( detector["DigitizerCard"]    )
  bl   = int( detector["DigitizerBlock"]   )
  ch   = int( detector["DigitizerChannel"] )

  card = cards.GetValueO(slot)
  helper = ConfigHelper(card)

  # -- digitizer block settings
  block = helper.block[bl]
  block.fir_enable = 'on'
  block.gate_window_length_ns = 2000
  block.sample_length_ns = 2000
  block.sample_start_ns = 0
  block.pre_trigger_delay_ns = 1000
  block.rise_time_ns = 16

  # -- digitizer channel settings
  channel = helper.channel[ch]
  channel.trigger_config = ['invert','internal']
  channel.fir_enable = 'on'
  channel.fir_cfd = '50-percent'
  channel.rise_time_ns = 16
  channel.gap_time_ns = 0
  channel.termination = '50Ohm'
```

Figure 5.33: A comparison of configuration interfaces.

allow the user to set and read values in human units, such as nanoseconds, while converting to machine units like samples behind the scenes. Named option properties allow the user to choose values for a digitizer option from a set of valid candidates. Finally, named bitmask properties allow the user to toggle boolean options using human readable names instead of flipping bits with binary math.

All of the property types mentioned above feature some form of input validation. For example, if the user attempts to set a 12 bit register to a 13 bit value, ConfigHelper will throw an error. This is in stark contrast to a plain NGM configuration, where this mistake would result in the 13th bit of the value being sliced off and the other 12 bits being written out of context. In addition to improving readability and error checking, ConfigHelper reorganizes the options from the NGM digitizer class. Instead of per option arrays indexed by channels or blocks, ConfigHelper represents each system, block, and card as individual objects with configurable properties. It does this by instantiating subclasses for the system, block, and channel option sets.

5.5.5 GUI

An important capability for the DDLI PoC was the ability to monitor the data in real time. Real time monitoring allows an operator to determine if all systems are functioning correctly during a run. It also gives the operator a tool for doing system troubleshooting or calibration. To facilitate such capabilities, a real-time GUI was developed using PyROOT.

The GUI was capable of displaying information from the last N seconds of data in a run, where N was a quantity controlled by the user. This window of data was grabbed from the collected data by a custom NGM module called TreeSplitter. By modifying the way in which files were saved by the NGM software, the dataset could be refreshed by the user at any time. Within the GUI, the user could select any digitizer card and channel from which to obtain data. The user could then select different methods of processing the data, mostly focusing on plotting energy spectra, but also including an overrange detection graph.

The DDLI GUI was integrated with the DetectorMapping information discussed in Section 5.5.4. This allowed for metadata about the detector events to be obtained, such as the type of detector they came from. With this information in hand, appropriate energy spectra could be plotted using the correct number of samples and with the correct spectral range. In the case of the dual-ended liquid scintillators, special processing routines were needed to combine the signal from both PMTs.

In addition to plotting histograms, the GUI was capable of smoothing them and plotting their derivatives. These two functions were useful for finding spectral features when combined with the included peak detection capability leveraging ROOT's TSpectrum class. With these powers, the GUI was capable of generating a full set of energy calibrations for all detectors in the system given a Cs-137 measurement. In the case of NaI(Tl), the full energy peak was used. For organic scintillators, the inflection point of the Compton slope was located and assigned the edge energy.

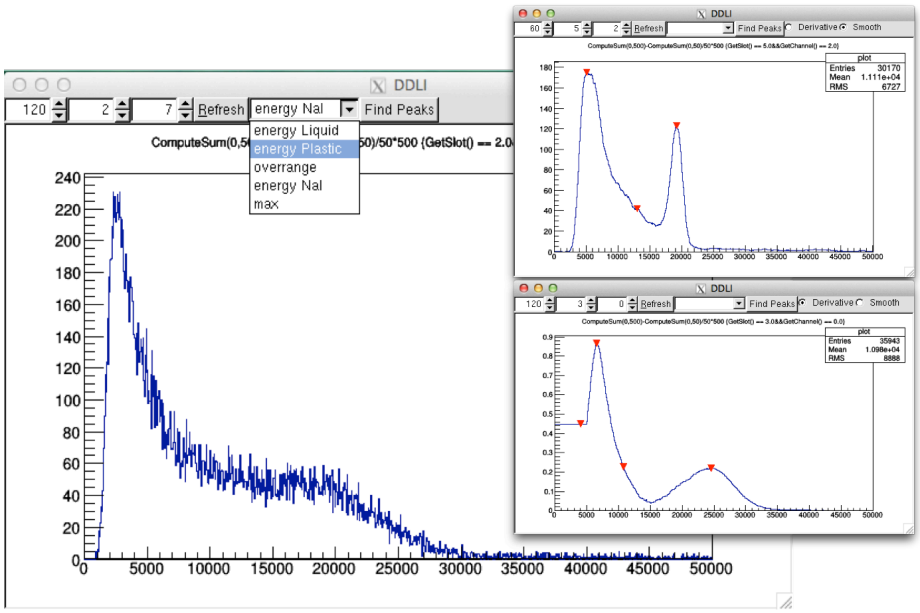


Figure 5.34: The DDLI GUI displaying live data from several channels.

It was important that the GUI be responsive, or else its usefulness would be diminished. Early versions took a considerable amount of time to generate graphs due to the shear volume of data being generated by the digitizers. To alleviate these problems, several optimizations were made. The greatest increase in speed was obtained by establishing a scheme to fill the digitizer scalars with the information necessary for baseline subtraction and pulse integration. Since the integration to the scalars is done by the digitizer FPGA, the waveform integration burden is shifted off of the GUI. The scalars also allow for the waveform buffer to be skipped during data reads. This is the main cause of speedup as the graphing process is memory bound. Further optimizations included things like plot caching to avoid repeated computation when switching between the various plot options.

5.5.6 Gain Matching

Gain matching was an important task in configuring the DDLI PoC trailer. The goal of the gain matching procedure is to adjust the bias on each detector in the system to equalize their response. Response in this case refers to the integrated charge output

for a given energy deposition in the detector. Gain matching makes analysis of the data easier, since the threshold and efficiency differences from detector to detector will be minimized. One might think that detectors of the same type would already give similar output for a given bias, but this is not always the case. Due to the compounded effects of varying photomultiplier multiplication, optical coupling, material batches, and other issues, significantly different levels of bias may be needed to equalize the output across a detector population, even when they share model numbers.

The DDLI PoC system contained three sets of detectors: plastic scintillators, dual-ended liquid scintillators, and an array of NaI(Tl) bars. Do to differences in their operation, each set of detectors was gain matched according to a separate procedure. The following sections will describe the gain matching process for each group.

5.5.6.1 Plastic Scintillators

As mentioned in Section 5.5.3.2, the plastic scintillator gain matching effort began with a great deal troubleshooting to enable the potentiometers in each unit. In the DDLI PoC trailer, a pair of detectors would share a single VME high voltage channel. To equalize the pair, the bias for one of the two detectors would be fine tuned by hand using a physical pot. The voltage for the pair's channel would be set in software.

To pair up the plastic scintillators, the detectors were all biased to the same voltage. A measurement was made using a set of Cs-137 sources placed along the length of the detector. From this measurement, the location of the Cs-137 Compton edge was found for each detector using the GUI tool. Due to differences in charge amplification, signal amplitudes varied from detector to detector. In order to place detectors with similar charge amplification in a pair, the plastic scintillators were sorted by the spectral channels of their Compton edges. The detectors were then rearranged inside the trailer to match this order.

With the detectors now in pairs, the gain matching could proceed. Cs-137 spectra were taken for all detectors at equal bias voltages. The bias voltage for these measurements was incrementally increased until the maximum operating voltage of

the detectors was reached. With all of these measurements in hand, the location of the Compton edge was found for each detector at each energy. This generated a gain curve for each detector which could be interpolated to find the bias which placed the Cs-137 Compton edge in a specific channel. Using these gain curves, a bias was selected for each pair based on the detector with the lowest signal amplitude. To maximize our PSD performance, the highest set of bias voltages possible had been chosen.

The reason the low-amplitude detector was selected to set the software bias involves the nature of the potentiometer. A potentiometer is only capable of reducing the bias seen by the detector. Therefore, gain matching with the pot could only be accomplished by reducing the voltage applied to the high-amplitude plastic scintillator. Once the voltage was set for all pairs, the GUI was used to monitor the location of the Cs-137 Compton edge as the pot was manipulated manually during an active measurement. The pot was adjusted until the pair was gain matched within a small margin of error.

Gain matching in this manner was time consuming, largely because the potentiometer installed by the manufacturer was extremely sensitive. It caused a factor of 10 change in signal amplitude over its full three-quarter turn range. Nevertheless, this procedure was painstakingly applied to each pair of detectors in series until the gain matching process was complete.

5.5.6.2 Dual-Ended Liquid Scintillators

The liquid scintillators in the DDLI PoC have two photomultiplier tubes per volume, one on each end. Each PMT on the detector has a dedicated VME high voltage channel, meaning that all gain matching can be done in software. What complicates the gain matching procedure in the liquid scintillators is the fact that signal amplitude for an event is dependent upon the position of interaction. Because events can occur along the entire range of the liquid volume, some events will occur closer to a given

PMT than others. An event occurring close to a PMT will appear brighter than an event depositing the same energy further away. This has the effect of blurring out the Compton edge to an extreme degree when looking at a single PMT.

To overcome the washout from position dependence, the position dependence itself was exploited to select only events occurring near a given PMT. This was accomplished by recognizing that the ratio of amplitudes from the two PMTs for a single event encodes the position along the length of the detector. By placing a Cs-137 source level with the center of the detector and selecting events in the top few percent of a contrast metric, we were able to select events within a short distance of the PMT under consideration. Spectra from this subset of events were much improved.

Using this technique, a series of Cs-137 measurements were conducted over a range of detector biases. This measured data was analyzed in the same manner as for the plastic scintillators. In this case however, the gain could be set in software for all the channels under consideration. In operation, the individual PMT signals would be combined to determine the deposited energy. This was accomplished by way of the geometric mean. Gain matching the detectors reduces the position dependence of the geometric mean of the two PMTs. Figure 5.35 shows the relatively flat geometric mean value of the Cs-137 Compton edge after gain matching as contrast (interaction position) varies.

5.5.6.3 Sodium Iodides

Unlike the liquid and plastic scintillators, which were connected to VME power supplies, the bias for the NaI(Tl) detectors was provided by a freestanding CAEN high voltage unit. This unit was used because the NaI(Tl) PMTs required positive high voltage whereas the organics needed negative. Since the freestanding positive HV unit was not connected the VME system, it could not be controlled through the NGM software. Instead, voltages were set via a Java web interface. This interface, while intuitive, did not allow for batch control of the detectors. All NaI(Tl) voltages would need to be entered by hand.

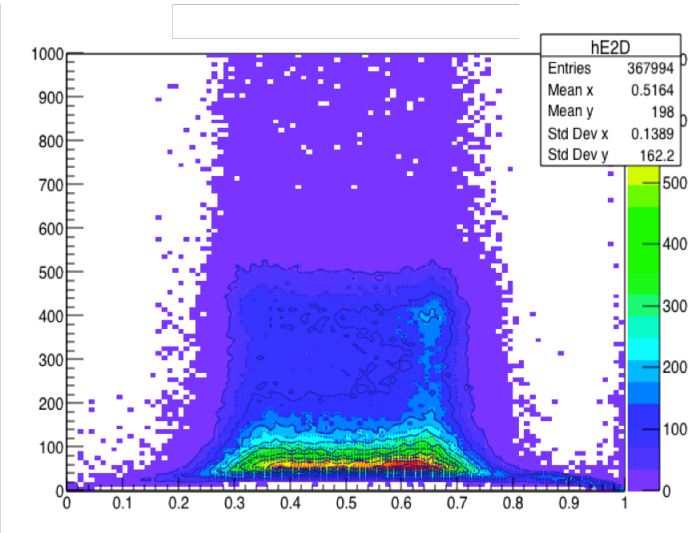


Figure 5.35: A 2D histogram showing the geometric mean of PMTs in a dual-ended liquid scintillator vs contrast for Cs-137. The relatively flat shape of the Compton edge feature indicates that there is low position variation in the geometric mean as a result of gain matching.

To get the high voltage into the right ballpark, data was taken for the NaI(Tl) detectors at a range of voltages. Since this work could not be automated and was fairly time consuming, the voltage step size across the operational range was larger than for the organic scintillators. This data was then used to solve for a voltage which would bring the detectors into a nearly gain matched state.

Following this procedure, the detectors were then actively gain matched during a measurement by modifying their voltage in the web interface. The location of the Cs-137 full energy peak was monitored as the voltage was updated. Over the course of a few days, all NaI(Tl) detectors were gain matched to within approximately 1 volt of their ideal bias.

5.5.7 Navigation

The DDLI PoC was equipped with several navigational systems. These would be used to keep track of the trailers position and displacement during a measurement, as well as ensure that the data stream was aligned to a reliable clock. In the following sections,

the installation, configuration and operation of three systems will be described. The first, a GPS unit, was used to tie system timing to satellite time. The second, a Doppler "fifth wheel", was used to monitor the speed and displacement of the ddli system. Finally, the differential GPS system was used to provide a centimeter accurate position for the trailer during controlled measurements.

5.5.7.1 GPS

The GPS receiver in the DDLI PoC trailer was capable of providing position information for the system if it had been desired. More importantly though, was its ability to sync our electronic clocks with satellite time. Satellite time is the clock used by GPS satellites. It is an atomic clock with a fixed offset from UTC. Instead of allowing the less reliable internal clocks on the VME crate or system computer to assign time to our data, it was desirable to use a standardized time. Not only would this give us better timing within the DDLI system, it would allow the DDLI data stream to be synced with any other UTC system.

The systems GPS receiver could make its satellite clock available in two ways. The first was through a computer network as an NTP (Network Time Protocol) server. An NTP server is used to constantly synchronize time between multiple computer systems. Software on the client computer tracks messages from the NTP server to know the current time. The client software also attempts to correct for inaccuracies induced by network latency and variability.

The GPS receiver was connected to the DDLI PoC's embedded computer via a router inside the trailer. NTP service on the GPS receiver was enabled through a front panel interface. A static IP address could be selected for the NTP server in this manner as well. The computer was then configured to use the GPS box as its NTP server. Best results were obtained by using the shortest polling time in the clients NTP configuration. After some trial and error, the *ntpstat* program reported alignment to within 5 ms, which was good enough for our application.

The second way to take advantage of the GPS clock was through a 1 pulse-per-second logic output. This logic pulse was produced when the GPS unit received the second-boundary pulse from the satellites it was tracking with the roof mounted GPS antenna. The 1 PPS signal was sent directly to a digitizer card for inclusion in the data stream. Because the digitizers timestamp events using their own internal clock, which runs independently of the NTP controlled computer, the 1 PPS signal was needed to anchor the VME time to UTC.

5.5.7.2 Doppler 5th Wheel

The Doppler 5th wheel is a small device which was attached to the underside of the DDLI PoC trailer. It was originally used for the Large Area Imager. The 5th wheel produced a pulse at a set distance interval as the trailer moved. This gave the system a very accurate odometer.

To test the unit, it was first removed from the underside of the trailer, where it was still mounted. As the wire had been run through the floor and sealed, the fifth wheel wiring needed to be cut. The unit was taken inside and wired to a laptop charger brick to provide power. The signal wires were then connected to a BNC cable. Operation was confirmed by monitoring an oscilloscope while waving a hand in front of the device.

To remount the fifth wheel on the trailer, connectors needed to be affixed to the ends of the cut wires to reattach them. The fifth wheel can be seen during the rewiring process in Figure 5.36. Inside the trailer, a solder-free connection box was used to attach the laptop charger and signal cables to the device via a wire through the floor. The power brick was plugged into the trailer's internal power, while the signal was piped into the digitizer.

The exact shape of these pulses was dependent upon the angle at which the device was pointed toward the road and the speed at which objects were approaching. Configuring the digitizer to trigger on these pulses proved somewhat difficult.

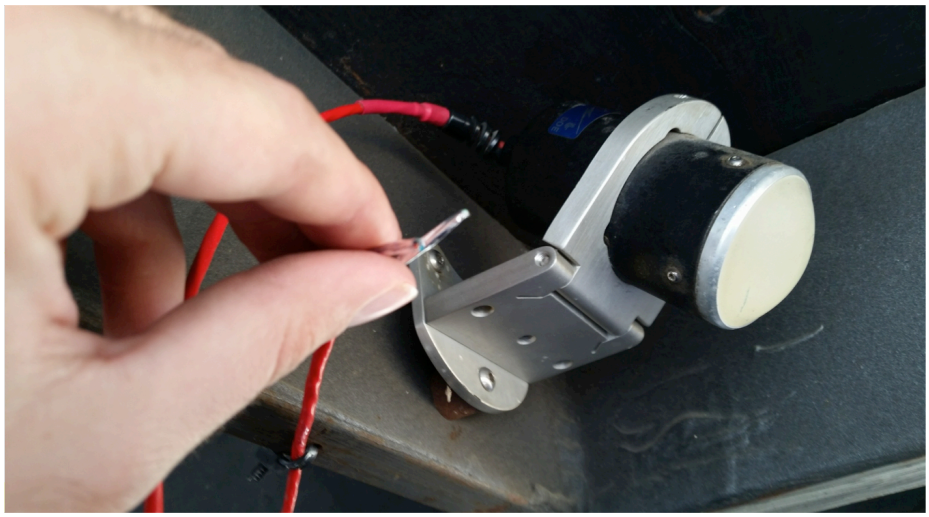


Figure 5.36: The Doppler 5th Wheel during reinstallation.

Eventually however, fifth wheel pulses from hand waving were being faithfully recorded by the digitizer. A more rigorous test would have to wait until the trailer was ready to drive.

5.5.7.3 Differential GPS

Normal GPS is capable of determining position within a range of around 10 meters. To obtain centimeter-level position accuracy, a Differential GPS system can be used. Differential GPS provides position relative to a fixed location by using more than one GPS receiver. As it turns out, atmospheric conditions are one of the main causes of error in GPS positioning. If GPS receivers are near one another, the signals reaching them will experience similar conditions along the way. This results in a highly correlated error in the position reported by nearby GPS receivers. This correlation means that the displacement calculated by subtracting receiver positions can be reported with high accuracy. Furthermore, if one of the GPS units is placed at a known GPS coordinate, accurate absolute positions can also be obtained.

Considering these benefits over normal GPS positioning, a DGPS system was used during the DDLI PoC measurement campaign. The DGPS units were from the ProMark series manufactured by Ashtech. The DDLI PoC setup used two receivers,

a stationary base unit near the measurement site and a rover unit inside the trailer. The base unit was connected to a GPS antenna on a tripod while the rover unit was connected to a GPS antenna mounted to the roof of the trailer.

The DGPS units were operated in a post processing mode. This meant that each unit was started before the measurement and allowed to collect data for the duration. This data was written in "G-file" format to an SD storage card in each device. After the measurement was complete, the G-files could be transferred to a computer and used together in a post processing stage. DGPS data was therefore saved and controlled independently of regular data acquisition

Post processing of the DGPS data was performed by the GNSS Solutions program. GNSS Solutions calculated the GPS coordinates and differential vectors for the receivers on each second boundary of the data. This information was then output to text files for later processing.

The first DGPS tests were performed by setting up the base unit and walking around with the rover in hand. Initially, it was difficult to get the receivers to acquire the satellites, but after flashing the memory and fixing a corrupted file structure, a connection was achieved. More trouble was encountered when post processing was attempted. The version of GNSS Solutions that had come with the DGPS units would not process the G-files. Surprisingly, the manufacturer was not able to help because they had been acquired by another company. Fortunately, Mark Silver, a DGPS expert at iGage Mapping Corporation, was able to help. Mark Silver forwarded an updated copy of GNSS Solutions which was able to process our data files. He also clarified several questions the author had about DGPS and the GNSS Solutions software.

5.5.8 Driving

With all the essential subsystems installed and operational, it was time to begin configuring the DDLI PoC as a whole. The following sections will describe the first mobile, full system tests.

5.5.8.1 Weight Distributing Hitch

Due to the DDLI systems heavy weight, it could not be legally towed using the trailer hitch from the Large Area Imager. Instead, a weight distributing hitch had to be installed. This installation process took a few days, after which time the trailer was towed by the author for the first time from the shop to the 3500 building at ORNL.

Before the trailer could be driven, the weight distributing hitch needed to be set up following a careful procedure. First, the hitch was attached to the truck. Next the weight distribution bars were inserted into the hitch. Chains from the bars were then fixed to the truck using two cams. A second set of chains was attached to the truck for backup. Next, a breakaway cable was connected which would cause the trailer brakes to engage in case of unintended separation from the truck. Finally, the brake light connection and a locking pin were put in place. A photo of the hitch can be seen in Figure 5.37.

5.5.8.2 First Drive

The DDLI trailer was ready to be towed. All that remained was to power up the generator. Up until now, DDLI PoC had been powered by a 30 amp line from the 3500 building at ORNL. Before we could do a mobile test, we would need to bring the generator online.

The first step was naturally to acquire fuel. While the DDLI PoC trailer runs on regular gasoline, ORNL employees are not allowed to work with this common

Figure 5.37: A photo of the weight distributing hitch.

substance without special training. Instead, ORNL operations needed to be involved to carry out the fueling. Before we could acquire fuel, the generator needed to be added to the ORNL database.

After obtaining fuel, the first attempts at starting the generator were unsuccessful. Unsurprisingly, the battery was dead after many months without use. A truck was used to jump the generator, allowing it to start. The generator stalled, however, as soon as the jumper cables were removed – we would need a new battery. Fortunately an identical battery was available from the parts store on the ORNL campus. To charge it, a battery charger was scavenged from the Roadside Tracker, another threat detection system at ORNL which was no longer in use. After charging the battery, the generator started easily.

With the generator taken care of, a first drive was planned. This would be a simple trip around the ORNL campus to determine if the fifth wheel, GPS, and detectors were all operating as expected under generator power. To start, all critical systems were moved from wall power to the UPSs (Uninterruptible Power Supplies). The UPS units in the trailer provide short term AC power from batteries in case of power loss. Next, the weight distributing hitch was attached to the towing truck. Next, the

generator was started. With all systems powered up, the trailer was disconnected from shore power. The switch inside the PoC trailer automatically handled the transition from shore to generator power. The DDLI system was ready to go.

At this point, data acquisition was started. Once the run start was confirmed, the trailer was towed for the first time while taking data. After a trip around the block, the trailer was parked to see if everything had worked. It appeared that all systems were in working order except for one: the fifth wheel. For some reason, no triggers had been accepted. To ascertain the problem, the author crawled underneath the truck and began testing with an oscilloscope. It was at this time that the generator began to fail.

5.5.8.3 Generator Troubleshooting

During the first driving test, the generator began sputtering and losing power. After it was shut down, it would not restart. It was initially speculated that a faulty fuel filter was the cause. However, the engine would still not start after a new fuel filter was installed. At this time, we began to suspect overheating. After reading the generator manual and searching online, however, it was discovered that the generator did not power its own battery in the way an automobile engine does. Given the typical use case for a generator of this type, this makes sense. This type of large generator is typically used infrequently or in power outages. In these cases, a battery only charged when the generator was running would be long dead. We would need to keep a battery charger attached to the battery at all times.

The battery charger from the Roadside Tracker was spliced to a long set of leads which were run through the trailer wall. These leads were then attached to the battery. The battery charger was then plugged into wall power inside the trailer and set to trickle charge. In this way, the generator could charge its own battery while it was running. When not running, shore power would take over and continue to keep the battery charged. To test this setup, the generator was run under load for an hour with no issues.

5.5.8.4 Second Drive

With the generator issues fixed, a second test drive was in order. The PoC trailer was hitched to the truck and powered on. Before driving, a Cs-137 measurement was taken for energy calibration. Next, the trailer was towed to Bethel Valley Road. The DGPS system was then set up, with he base unit placed on the side of the road and the rover connected to the trailer. Finally, several laps were driven around the ORNL campus.

During this process, the author was monitoring the system output using the GUI described in Section 5.5.5. At this time, a few issues stood out. First, it appeared that the gain matching performed some time back had not held. It would need to be repeated, at least for the NaI(Tl) detectors. Second, the lower thresholds for all of the detector channels were very high. This would need to be fixed in the configuration file. Finally, during the drive itself, several NaI(Tl) high voltage channels tripped by exceeding their amperage safety threshold and were automatically shut down. The cause for this was not clear.

Over the next few days, the gain matching procedure was repeated. Updates were also made to the peak finding routine to locate the lower threshold for the detectors. These thresholds were subsequently reduced as close to noise levels as possible on each channel. The NaI(Tl) detectors that were tripping were also examined. Some simply needed their maximum current draw level increased by a small margin. Others, however, were tripping at low voltage levels. In the latter detectors, a short was likely present. The faulty detectors were taken to the lab and outfitted with replacement PMT bases, as this was the most likely place for a short. After confirming that the problem was resolved, the detectors were mounted back in the trailer.

5.6 Mobile Measurements

This section will discuss the measurements performed with the DDLI PoC at Oak Ridge National Laboratory. Much of the work described here was conducted in parallel with the trailer build.

5.6.1 Measurement Goals

As discussed in the Appendix B, the DDLI PoC measurements were originally intended to tie back into the simulation effort. Performance reporting for the DDLI system was to be based on a source injection study with measured background and simulated sources. The measurements for the DDLI PoC were designed with this source injection effort in mind.

To understand the performance of the Large Area Imager (see Section 5.2.1.1), an injectable source was created not from simulations, but from real measured data. This was done by measuring a source and background long enough to obtain good statistics, then subtracting the background to leave only the source response behind. For the DDLI PoC, we would seek to do the same.

An injectable source from real data would be advantageous in several ways. First, it could be used as a direct comparison to a source-only simulation. By observing the difference in response between our simulated and measured test source, we could determine whether our model was behaving in a realistic way. With the addition of a background dataset, we could also perform source injection comparisons as a control. If our algorithms gave similar performance results both the injected and simulated test source, it would lend confidence to simulations with other, unmeasured source configurations. Source injection studies might also direct the tuning of Geant4 models. With alarm performance as a cost function, the need for simulation tweaks would be easier to evaluate.

With these benefits in mind, the following plan was developed. Our injectable source would be created by driving the DDLI PoC trailer past a series of sources

at extremely slow speeds, perhaps multiple times. These sources would include an unshielded gamma ray source and an unshielded neutron/gamma source. The neutron emitting source would be the most important, as our goal would be to determine the added benefit of neutron detection. Similar measurements without sources would also be conducted to measure background. Subtraction of this background from source runs would yield the injectable sources we were after.

It is important to clarify the measured quantities in these experiments. Our goal for each measurement would be to understand count rate as a function of position. As the trailer was towed past a source, DGPS data would be used to map counts into the world. These integrated counts would be divided by the time spent in any particular location to yield count rate. Count rate profiles for the measurement path would be built for each detector independently to enable coded aperture imaging. If spectroscopy was desired, per energy-band count rates would also be required.

When considering imaging, it is important to remember that the built PoC system was not fully instrumented. Only 5 out of 21 liquid scintillators were present and only one row of plastic scintillators had been installed in the mask. While the counting data from the plastic scintillators in the mask could likely be scaled up without much issue, the liquid scintillators were more of a problem. A pass by a source with 5 detectors would not be sufficient to create an imageable response.

To remedy this problem, it was decided to repeat measurements with shifts in the liquid scintillator positions. The five liquid scintillators would be installed as a group on one end of the array for the first measurement. When the measurement was complete, they would be shifted down five spaces and the measurement would be repeated. For a given source, four measurements would be taken, allowing 20 of the 21 liquid scintillator locations to be covered. The data from these four measurements could then be combined to simulate a single fully instrumented acquisition. Using this approach, it would be possible to produce an injectable source for neutron coded aperture with only the partially instrumented prototype.

5.6.2 Choosing a Location

Before measurements with the DDLI PoC could be taken, a measurement location would need to be selected. The selection of the measurement site was subject to several constraints. Because ORNL sources would be used, the measurement would need to take place somewhere on the ORNL campus. The site would need a straight and level stretch of drivable terrain bordering area clear of obstructions in which the source could be placed. Background variations within the area would need to be minimized, so road material changes and structures should be avoided. Additionally, to facilitate the generation of complete coded aperture images, a pass by the source should cover around three field-of-view lengths. Depending on the standoff distance, this could mean close to 1000 feet of drivable path. Finally, to minimize complications related to sources and extremely slow driving speeds, a low traffic, low population area was desirable.

To map candidate locations, a police rangefinder, shown in Figure 5.38, was used. Several candidate locations were evaluated, keeping the aforementioned constraints in mind. The maps for several of these locations can be seen in Figure 5.39. Eventually, a parking lot at the intersection of Bethel Valley Road and Spallation Drive was chosen. This location met all of the conditions. Bethel Valley Road provided over 900 feet of driving distance over consistent material. The adjoining parking lot formed a circuit which could be lapped without difficult turning. The empty field to the north would provide a clear line of sight to the source and significant standoff distance if desired. Bethel Valley Road and the parking lot were, however, high traffic areas. Fortunately, this problem could be mitigated by conducting measurements afte hours when activity on campus was reduced to nearly zero. A satellite image of the measurement location is shown in 5.40.

Figure 5.38: Laser rangefinder used to map and evaluate candidate measurement sites.

5.6.3 Choosing Sources

The highest activity sources available would be the best choice for our experiment in terms of statistics. Unfortunately several other factors were at play in our selection. One especially important factor was the degree of RCT involvement in source transport. Even though the measurement site at Bethel Valley Road was inside the ORNL campus, any radioactive source transported on Bethel Valley was subject to Department of Transportation rules. Above a certain activity, sources would need to be packaged in DOT compliant containers by trained technicians, transported by RCTs, and unpacked at the measurement site. When viewed in light of the small remaining budget for the project, these costs would not be acceptable.

To make source transport easier, we would need to pick sources which were below the limits for a DOT Type B quantity. ORNL's onsite transportation document declares sources in this activity range exempt from several packaging and transportation approval requirements. Sources below Type B would therefore be transportable to the measurement site at any time by an RCT.

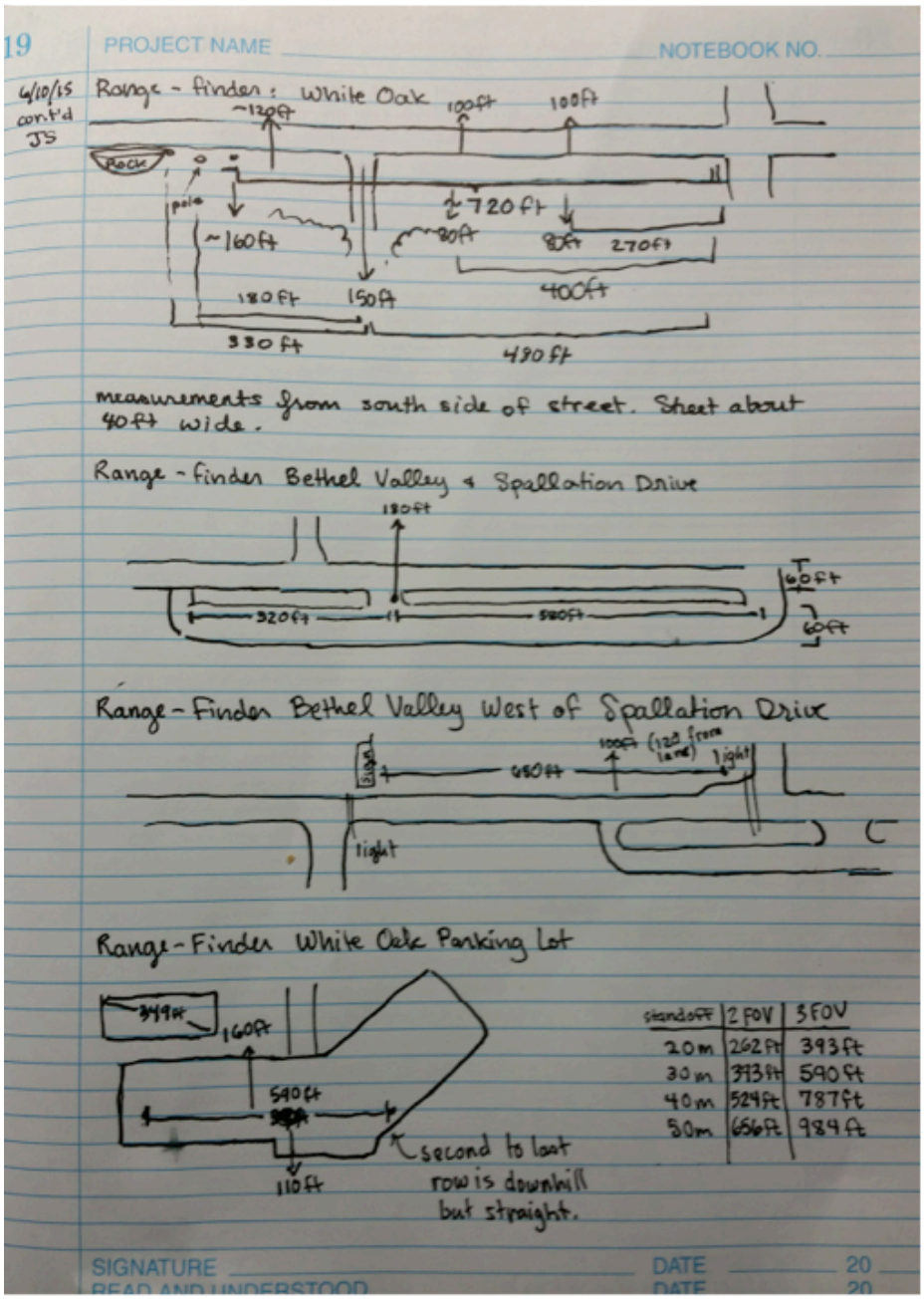


Figure 5.39: Hand drawn maps of several candidate sites.

Figure 5.40: Satellite view of selected measurement site on Bethel Valley Road. Red line marks the path the DDLI PoC trailer would take during measurements.

As it turned out, the highest activity Cf-252 source available in the NSITD group's inventory had decayed enough to meet these conditions. This source would provide our neutron signature. As an easily identifiable test source, a Cs-137 source was also selected. Shortly after our measurements were completed, the Cf-252 source was measured by another group of researchers using the Californium Shuffler. They found the source to emit neutrons at a rate of 279,000 per second into 4π. As the Californium Shuffler is very accurate, this indicated some error in the ORNL source database. Of course, we were not aware of the reduced source strength at the time of our measurements.

It should be noted that while these sources were easily transportable, they were not especially hot, especially the Cf-252 source. Many laps would be required to measure the source signature. Additionally, it would likely not be possible to see the Cf-252 source through its gamma ray signature, as the gamma ray background is considerably more intense. A back of the envelope calculation suggested that around 10 neutrons per second might interact from the source at 20 meter standoff. This was a very low rate, but at the time, this was the best available option for a neutron source.

5.6.4 Measurement Plan and Radiological Work Permit

Before work could begin, a measurement plan and associated radiological work permit would need to be approved. These documents are part of the safety culture at ORNL and help make sure measurements are conducted in an appropriate way. These documents would need to go through several revisions and get approval from supervising staff before they could go into effect.

The first draft of the measurement plan contained the general plan for measurements. The language was then updated to reflect DOT language by Gomez Wright. It was then passed to Dale Perkins for final adjustments. Next, the measurement plan was added as an attachment to our Research Safety Summary. The full text of the measurement plan can be found in Appendix E.

There are several key points in this measurement plan which have not been discussed up to this point. In order to maintain source control, two Radiological Workers (Rad Worker is an ORNL training qualification), would need to be present at all times. This was to enable at least one person to monitor the source at all times. Given the low remaining budget, this would be an important concern, as staff members come at a cost. In the same vein, an RCT and government vehicle would be needed to transport the sources to the measurement site. As the measurements would take place after hours, this would normally cause the project to incur recurring overtime costs for source transport. To avoid these costs, it was approved for sources to be stored in the DDLI PoC trailer when not in use under lock and key. Since only an RCT can transport the sources to and from the measurement site, the DDLI PoC trailer would need to remain parked at the measurement site when not in use.

With an official measurement plan in place, work on a Radiological Work Permit began. Debra Austin, the RCT for NSITD, drafted the RWP according to the measurement plan. As Radiological Workers, we would need to sign in on this RWP

whenever we were working with these sources. The RWP included information about our allowable doses, the dose rates from sources, source control, training requirements, and the protected area around the source.

5.6.5 Setting Up Sources

The first source measurements would be done with the Cf-252 source, as it was most important. As the day for source measurements drew closer, it became clear that there was an oversight in our measurement plan. As mentioned previously, the DDLI PoC trailer would need to remain parked at the measurement site whenever a radioactive source was present. This resulted from the combination of after hours measurements and the requirement that an RCT perform all source transportation. The oversight came in the form of electrical power.

With measurements set to take place in the middle of summer, the DDLI PoC trailer would need to keep its air conditioning running. While the trailer has a generator, it was not capable of running for 24 hours on a single tank of fuel. Even if it could, the cost of keeping the trailer fueled would be large, considering overtime pay for operations employees. While the trailer could be powered by a 30 amp line, we were not allowed to tow the trailer back to building 3500 with the sources inside.

Initially, we attempted to find building power within the 7000 block of buildings near our measurement site. If the trailer could be parked within walking distance of the measurements, we could walk the sources to lockup without violating any rules. After speaking with many building managers and their supervisors, it looked as though we could park the trailer outside of a rarely used building. Unfortunately, the only power available was attached to a 20 amp breaker.

We needed to find out whether the DDLI PoC trailer could run on a 20 amp circuit. A 20 amp circuit can only handle a continuous load of about 16 amps without tripping. With only the air conditioning running, we measured the current draw of the DDLI trailer to be 15.5 amps: uncomfortably close to the limit. If the 20 amp

circuit breaker tripped, we would be required to spend around two thousand dollars to upgrade the circuit to 30 amps. At this point in the project, we had neither the money nor time to get such work done.

Fortunately, there was another solution. As it turned out, the DOT inspection from the neutron background measurements trailer (see Section 5.3.2) was still in effect for another two months, meaning the DGNI trailer was still street legal. By towing the DGNI trailer to the measurement site, we could gain a perfectly compliant source storage locker. This would free the DDLI PoC from its obligation to remain onsite, allowing it to be towed back to building 3500's 30 amp power.

With this problem solved, the first source was transported to Bethel Valley Road. The Cf-252 source was brought to the site first, as the neutron measurements were the most important for the project. Only one source would be brought to the site at a time to avoid contaminating measurements with secondary sources. The NSITD's RCT, Deb Austin, transported the source to the site in a lead and polyethylene pig. She then set up a stanchion and radioactive material posting. A spot approximately 20 meters from the road and at the midway point of the path had been selected for the source location. The RCT performed a background survey followed by dose measurements with the source in place. A photo of this process can be seen in Figure 5.41. When the surveys were complete, the Cf-252 was stowed in the DGNI trailer behind lock and key. This process was repeated when the Cf-252 was later swapped for Cs-137.

5.6.6 Conducting Measurements

5.6.6.1 Background Measurements

The first mobile measurements with the DDLI PoC began in July, 2015. Because the problem of source storage was still being worked out, background measurements

Figure 5.41: The RCT Deb Austin performing a survey at the source location.

were conducted first. These measurements were completed over the course of two nights in early July by the author. A procedure for the remaining measurements was developed at this time.

To begin, the DDLI PoC trailer's weight distributing hitch was connected to the NSITD's Ford SuperDuty truck according to the procedure described in Section 5.5.8.1. The generator was powered up and the trailer was disconnected from shore power. Next, the data acquisition system, GPS box, fifth wheel, and high voltage were powered up. Six Cs-137 sources were placed along the length of the trailer interior for an energy calibration measurement. During this measurement, the DDLI GUI was used to confirm that all detectors were operating as expected. After the measurement, the Cs-137 sources were returned to the lab in building 3500. At this point, the trailer was ready to be towed to the measurement site.

Once on location at Bethel Valley Road, the trailer was parked in the adjoining parking lot. The DGPS base unit and tripod were taken to a set position in one corner of the parking lot. A GPS antenna was attached to the tripod and connected to the DGPS base. The antenna height was measured and used to configure the base

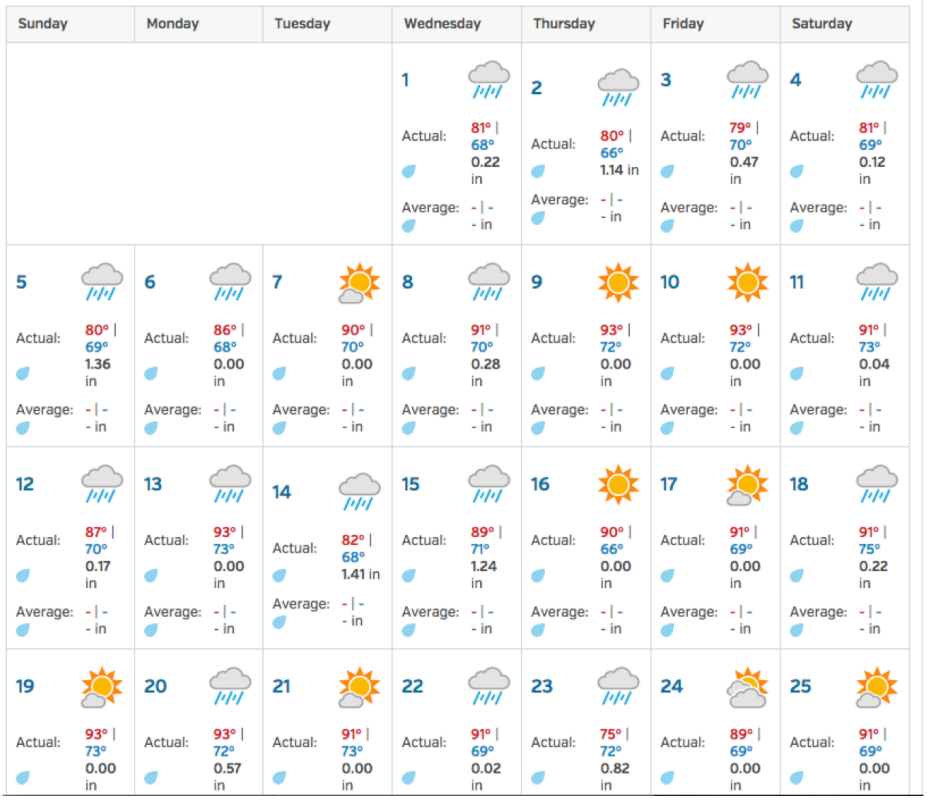
unit for measurement. The base measurement was then started for the evening. Back in the trailer, this configuration was repeated for the rover unit after connecting it to the roof mounted antenna. The DGPS system was now up and running.

All preparations were now complete for the measurement. Using the trailer's wifi network, the high voltage was enabled and the data acquisition system was started. As a test the fifth wheel, the trailer was towed forward for a few feet. Once triggers were registered for this movement, the measurement began in earnest.

The trailer was towed out onto Bethel Valley Road. Once out of the turn, the speed was decreased to approximately 1 mile per hour. This speed was maintained along Bethel Valley Road until turning back into the parking lot at the end of the measurement path. Figure 5.40 shows the circuit driven by the DDLI PoC during the measurements. Ten laps around this circuit constituted one measurement. This took just over an hour to complete on average.

At the end of each set of ten laps, the liquid scintillators needed to be relocated. As mentioned in Section 5.6.1, this was necessary to perform coded aperture with our partially instrumented prototype. After parking the trailer, data acquisition was stopped and the high voltage for organics was disabled. Inside the trailer, the mounting screws, signal cables and high voltage lines were disconnected from each of the ten PMTs. The five liquid scintillators, now free, were shifted exactly five spaces in preparation for the next measurement. The PMTs were then rewired and the mounting screws put back in place.

Back in the truck, the high voltage for the organic scintillators was reenabled over wifi and another data acquisition was started. The measurement process was repeated and another ten laps were made. Ideally, this process would have been repeated two more times for a total of forty laps. Mother nature, however, had other plans. While shifting the liquid scintillators for a second time, it began to rain heavily. Due to the associated gamma ray background transient [45], this rain would put an end to background measurements for the night.



Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
			1 Actual: 81° | 68° 0.22 in Average: - | - - in	2 Actual: 80° | 66° 1.14 in Average: - | - - in	3 Actual: 79° | 70° 0.47 in Average: - | - - in	4 Actual: 81° | 69° 0.12 in Average: - | - - in
5 Actual: 80° | 69° 1.36 in Average: - | - - in	6 Actual: 86° | 68° 0.00 in Average: - | - - in	7 Actual: 90° | 70° 0.00 in Average: - | - - in	8 Actual: 91° | 70° 0.28 in Average: - | - - in	9 Actual: 93° | 72° 0.00 in Average: - | - - in	10 Actual: 93° | 72° 0.00 in Average: - | - - in	11 Actual: 91° | 73° 0.04 in Average: - | - - in
12 Actual: 87° | 70° 0.17 in Average: - | - - in	13 Actual: 93° | 73° 0.00 in Average: - | - - in	14 Actual: 82° | 68° 1.41 in Average: - | - - in	15 Actual: 89° | 71° 1.24 in Average: - | - - in	16 Actual: 90° | 66° 0.00 in Average: - | - - in	17 Actual: 91° | 69° 0.00 in Average: - | - - in	18 Actual: 91° | 75° 0.22 in Average: - | - - in
19 Actual: 93° | 73° 0.00 in	20 Actual: 93° | 72° 0.57 in	21 Actual: 91° | 73° 0.00 in	22 Actual: 91° | 69° 0.02 in	23 Actual: 75° | 72° 0.82 in	24 Actual: 89° | 69° 0.00 in	25 Actual: 91° | 69° 0.00 in

Figure 5.42: Observed weather for July 2015.

Rain would be one of the major obstacles in the DDLI measurement campaign. Figure 5.42 shows the recorded weather for Oak Ridge, TN during July 2015. Both gamma ray and neutron background levels are affected by rain storms. What is worse, these effects can last for several hours after a storm has completed. For this reason, even a small afternoon shower was enough to put a stop to measurements. [45]

The DGPS data acquisition was stopped and the base unit and tripod were loaded into the DDLI PoC trailer. The system was towed back to building 3500 and backed into its parking space. The hitch was then disconnected and the truck was parked. Before finishing up, the six Cs-137 sources were brought back down from the lab and placed in the trailer for a post measurement energy calibration. This measurement would help us diagnose any gain drift during the measurement. The remaining twenty background laps would be completed the following night by following the same procedure.

5.6.6.2 Source Measurements

Unlike the background measurements, the source measurements could not be performed alone. Per our measurement plan (Section 5.6.4), a two-man source control rule was in effect. Aaron Nowack and Micah Folsom, two graduate students from the research group, were kind enough to volunteer their time for two nights of measurements. This was sufficient to complete all forty laps of the Cf-252 measurements. Dr. Jason Hayward helped perform the Cs-137 measurements on a separate night.

The source measurements generally followed the same procedure listed for the background with a few extra steps. Once at the measurement site, an extra DGPS measurement was made at the source stanchion (Section 5.6.5). With the base DGPS in place, the rover unit was attached to the source stanchion. In this configuration, a static dwell was taken. This extra measurement would provide an accurate location for the source which would be useful in analysis of the data. Because the base location was never changed, we would now be able to determine the location of the imager relative to the source at any time.

Following this position measurement, the source was removed from the DGNI trailer. Recall that the DGNI trailer was being used as our onsite source locker. The source was then walked across the road to the source location in the north field and attached to the stanchion. From this point forward, it would be the job of the volunteer "second man" to watch the source until measurements were complete. Fortunately, WiFi was available from the nearby 7000 block buildings.

Over the course of three nights, sixty more laps were completed with the system. A full set of forty laps (four liquid scintillator positions) was finished for the Cf-252 measurements. Twenty laps (two positions) were completed the Cs-137 source. While not a complete set for Cs-137, it would be sufficient for troubleshooting.

Figure 5.43: The DDLI PoC trailer being towed during a source measurement. Pictured are the author (driving), and Micah Folsom, Aaron Nowack, and Jason Hayward (from left to right).

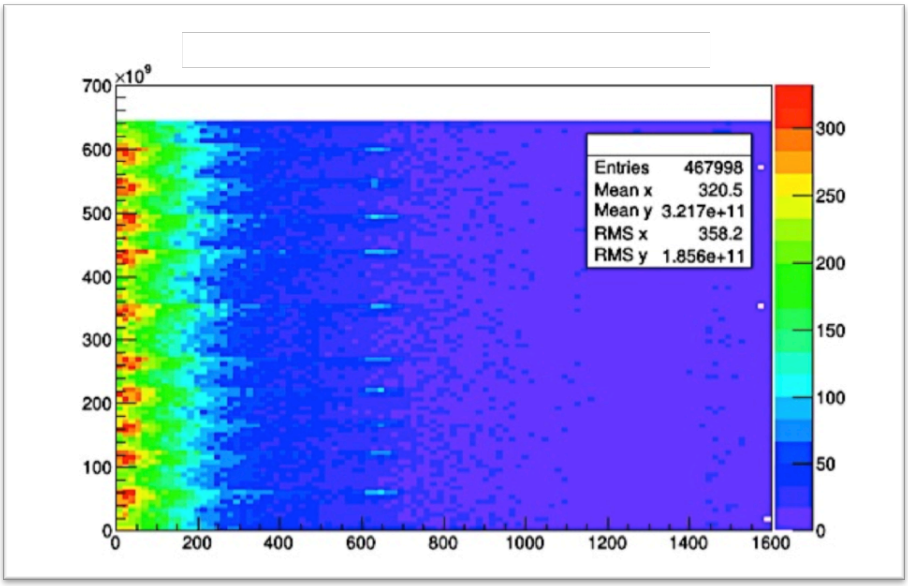


Figure 5.44: A histogram of NaI(Tl) events showing VME clock vs. energy in keV with counts represented by the color. Ten bright spots at 662 keV correspond to ten laps past a Cs-137 source.

5.7 Results

5.7.1 Analysis

5.7.1.1 First Look

With the measurements complete, the project moved into the data analysis phase. At this point, none of the measurement data had been analyzed. There had simply not been time to concurrently analyze and execute the measurements. Naturally, there was some worry that there might be problems with the data. As a first look, ten laps of Cs-137 data from NaI was imported and converted from the NGM binary format to a ROOT file format. A 2D histogram of this data was created with energy bins on the X axis and the the VME clock on the Y axis. This histogram is shown in Figure 5.44.

As hoped, ten distinct bright spots can be seen corresponding to the 662 keV Cs-137 gamma ray, one spot for each lap past the source. This was a good indication that

our data acquisition had been working as expected. To find out for sure, a significant amount of work would still need to be done. The final post processing goal would be to accumulate neutron counts into positions along the measurement path. In its current state, however, the data was not connected to GPS timing, DGPS position, energy calibrations, or any particle identity metric, all of which were required to do the count mapping. More experiments and new software would need to be written to incorporate to incorporate these pieces of data.

5.7.1.2 Event Timing and Position

In order to map detector events to points in space, we would first need to process our position information. As mentioned before, the DGPS system outputs G-files which can be processed in the GNSS solutions program to obtain centimeter accurate displacements. These displacements, called vectors, can be exported to text files along with GPS points. A Python script, *gpsToROOT*, was then used to convert these text files to the ROOT file format. During the conversion process, some geometry was used to relocate the origin to the source position. This gave us a DGPS ROOT file with a 1 Hz position of the trailer relative to the source. This process is visualized in Figure 5.45.

With the DGPS root file in hand, the next step was to correct the timing associated with each detetor event. Specifically, we would be converting the event's raw VME clock into a UTC time. This was done by combining the information from the GPS NTP, GPS 1 pulse-per-second signal, and the event timestamp. As described in Section 5.5.7.1, the DDLI PoC's embedded computer was synced to UTC time through the GPS NTP server. This meant that the starting UTC timestamp in the data stream was fairly accurate. Once the measurement began, the 1 PPS signal from the GPS would be used to increment the measurement's UTC clock. To make sure we knew the UTC time of the first GPS pulse, data acquisition was always started on a half second boundary.

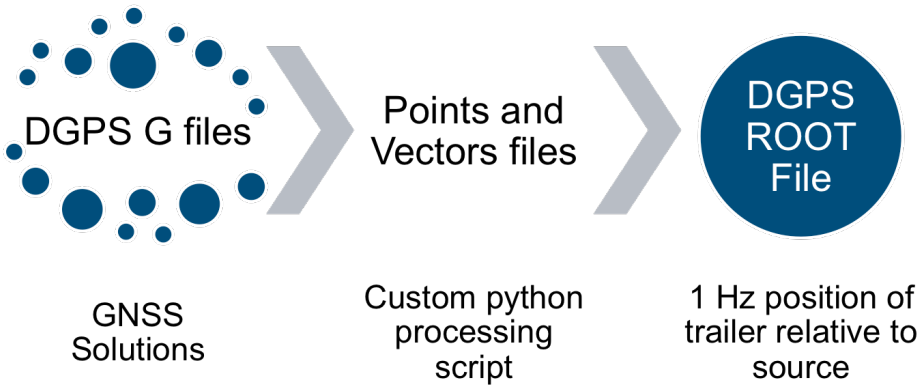


Figure 5.45: Diagram of the conversion process for DGPS files. G-files were translated into ROOT files which could be used throughout the post processing chain.

Analysis of the timing correction was less straightforward than one might think. This complexity arose from the unreliability of the GPS signal. While we have referred to it as a 1 pulse per second signal, in reality this signal arrived sporadically. Though it always fell on a second boundary, multiple seconds could pass between pulses. This was likely due to the age of the GPS unit. It was specified in the unit's manual specified that the timing circuit might need maintenance after a period of ten years, a point long past.

The conversion from VME clock to UTC was performed by an NGM module called *DDLITimeAdjustment*. *DDLITimeAdjustment* worked by estimating a conversion factor using a moving average. The sporadic nature of the GPS timing signal was accounted for by assuming that a VME clock second was at least reasonably close to a UTC second. With a conversion in place, each event could be assigned a UTC timestamp.

The DDLI PoC data stream now contained events with UTC timestamps and a DGPS position for the trailer on every GPS second boundary. With this information, it was possible to interpolate a unique position for every single detector event in a run. This process was carried out by the *DDLIPositionOut* module. A diagram

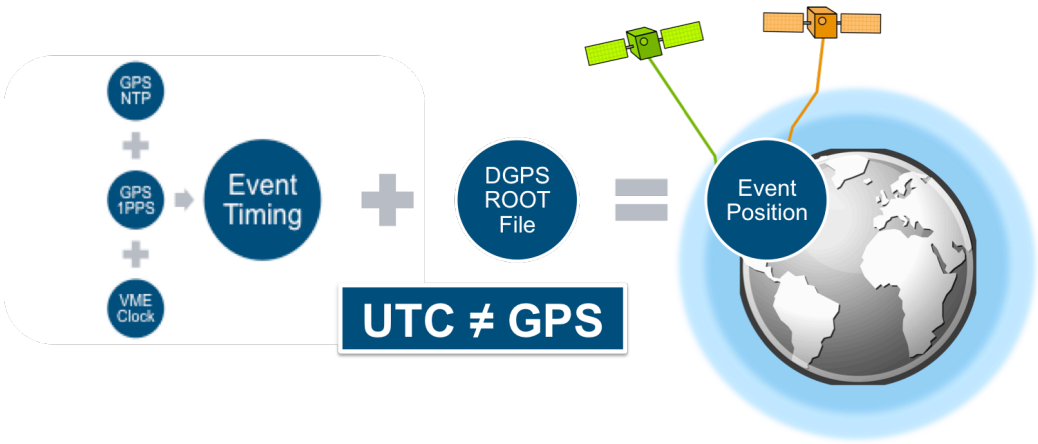


Figure 5.46: Diagram for position assignment to individual events. Event timing was used to interpolate positions from the DGPS file.

representing this process is shown in Figure 5.46. Particular care needed to be paid to timing standards at this stage. For example, UTC and GPS time are separated by 17 leap seconds. The NTP server accounts for these leap seconds; the DGPS system does not.

5.7.1.3 Energy Calibrations

The next step was to incorporate the energy calibration measurements we had performed before each night of driving. This energy calibration process was done automatically using the *ddli_energy_cal* Python script. It leveraged the same smoothing, peak, and derivative finding routines as the GUI. It also pulled in the DetectorMapping database to know which type of processing to perform on which channel. NaI energy calibration was performed using the Cs-137 full energy peak. The plastic scintillators were calibrated by assuming the inflection point of the Compton slope was at the Compton edge energy. The dual-ended liquid scintillators were calibrated using the Compton slope inflection point from the geometric mean of both PMTs. All calibrations were linear and calculated assuming a zero intercept due to baseline subtraction.

The output of the *ddli_energy_cal* script was a energy calibration Python Pickle file. This file contained a lookup table of energy calibrations for each channel. The Pickle file format was useful because it could be used when configuring the NGM library for post processing from Python. The energy calibration was applied by running the *apply_calibration* Python script. This script configured a *DDLIApplyCal* module within the NGM post processing chain, which converted each waveform integration to energy units via *NGMHitProcess*.

5.7.1.4 Count Rate Mapping

With this amount of processing complete, we could finally generate some mappings of our measured data. Counts were histogrammed by their 2D cartesian coordinates relative to the source position. This generated a map of all the counts from a measurement. A second histogram was generated based on millisecond counts from the measurement clock. This mapped the amount of time spent in each location. By dividing these two histograms, a count rate map was obtained.

Figure 5.47 shows the result of this operation for ten laps of Cs-137 data taken on July 17, 2015. This image was generated by gating on the Cs-137 full energy peak. There is a clear increase in count rate as the imager path approaches the source (located at the origin). This was a very good result and confirmed that the DGPS, timing, and post processing had all worked as planned.

A similar plot for the Cf-252 source data, shown in Figure 5.48, is less striking. This plot was generated using all detector counts without PSD. As mentioned in Section 5.6.3, the activity of this source was very low. It would generate approximately 20 gamma ray counts per second across the entire system compared to thousands from background. We would not be able to see the Cf-252 on the count rate map unless we were looking at neutrons. For this we would need pulse shape discrimination [46].

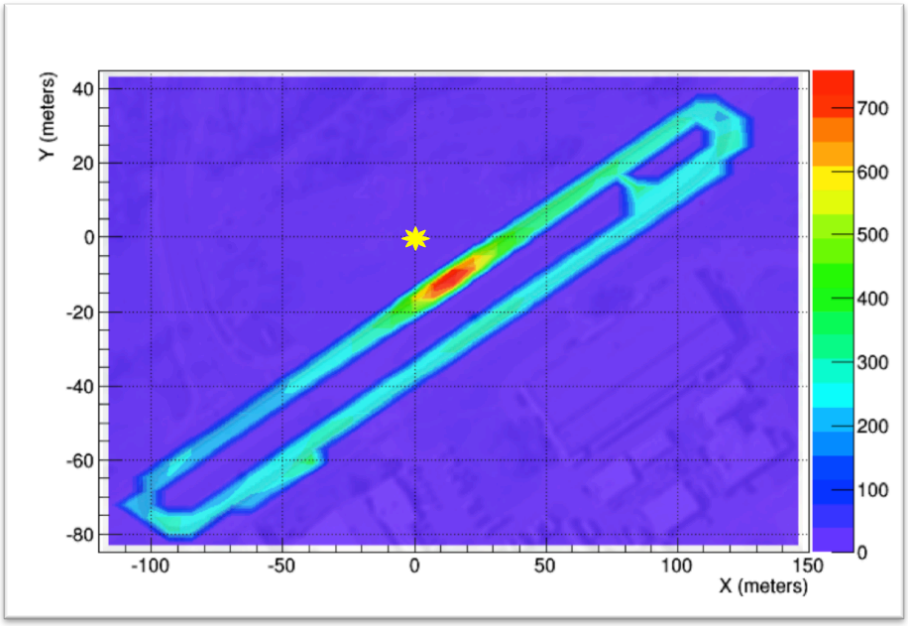


Figure 5.47: Count rate map for the Cs-137 measurements. This map was generated using counts under the full energy peak in the NaI(Tl) detectors. The X and Y axes represent distance from the source in meters. The colorbar represents the count rate in counts per second.

5.7.1.5 Pulse Shape Discrimination

To determine pulse shape parameters, a measurement would be needed. After a Cs-137 energy calibration, five Cf-252 capsules were brought into the trailer and arranged along its length. Data acquisition was then started for the organic scintillators and continued until there were at least one million events per detector.

PSD parameters were calculated using a modified version of NGMPSDMaker, a class originally written by Dr. Jason Newby. The class was augmented by the author to include extra rules for processing the double ended liquid scintillators. It was also modified to make it configurable from an independent script.

The PSD calculation was a two step process. First, the detector information was input into the NGMPSDMaker class using the DetectorMapping tool. This included energy calibrations and PSD gates which had been optimized for the best

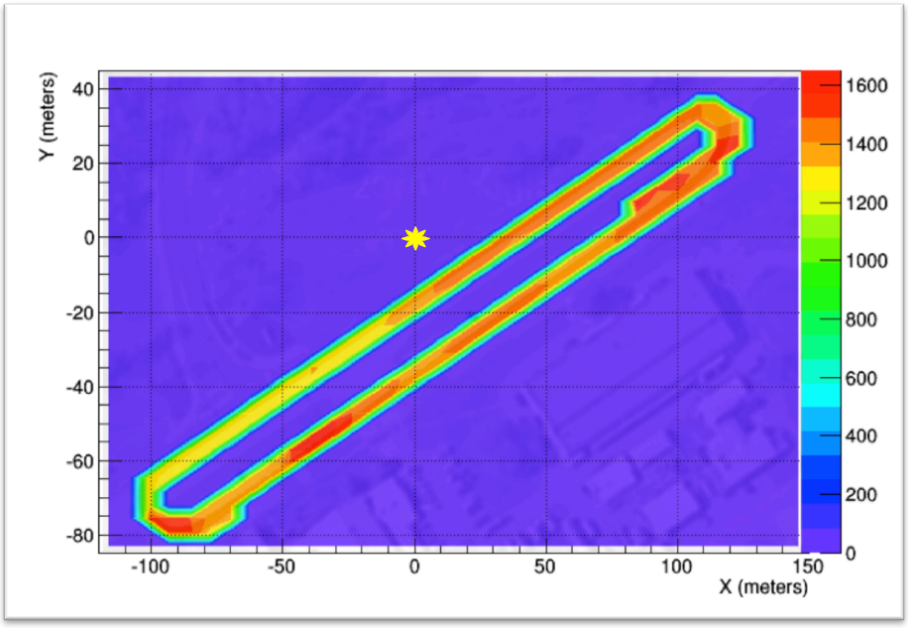


Figure 5.48: Count rate map for the Cf-252 measurement. This map was generated using all counts in the organic scintillators without PSD. The X and Y axes represent distance from the source in meters. The colorbar represents the count rate in counts per second. The source is not detectable due to the high gamma background.

neutron/gamma separation. After the PSDMaker was configured, it processed the Cf-252 measurement data to generate PSD plots like the one shown in Figure 5.49. These were saved as histograms in a ROOT file.

In the second step, the ROOT file of histograms was reopened. These plots were then processed by NGMPSDMaker to fit the neutron and gamma bands. Cubic splines were used to describe the mean and standard deviation of the neutron and gamma bands as a function of energy. These splines were output as NGMHitProcess objects and saved into a separate ROOT file. With this PSD ROOT file, it was possible to assign an identity to each event from an energy calibrated measurement with the DDLI PoC.

After processing the Cf-252 data from Bethel Valley Road to include PSD, it was possible to select neutron events only. To avoid gamma contamination, a 5σ cut was

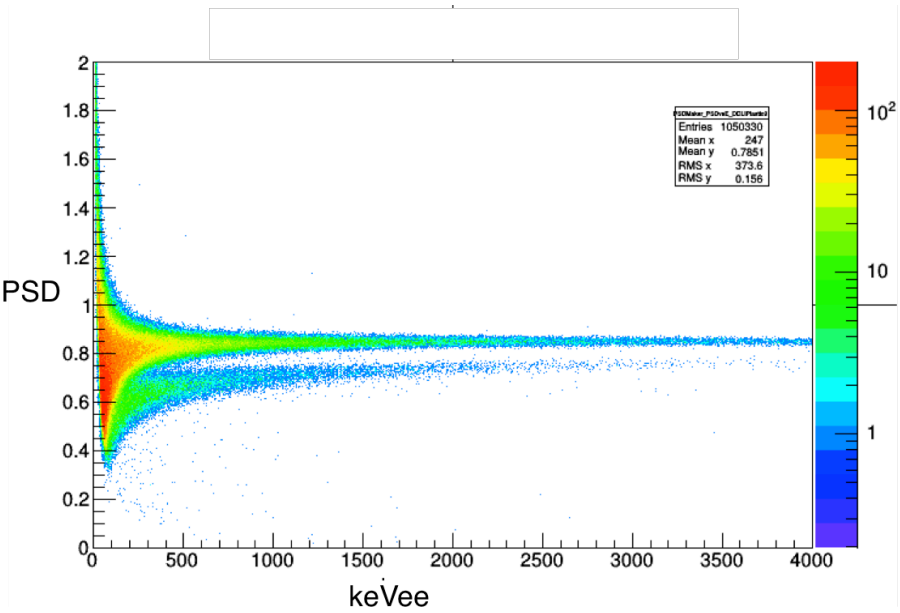


Figure 5.49: Example of a PSD plot generated by NGMPSDMaker.

placed on the gamma ray band. The resulting count rates were mapped into 2D space to generate Figure 5.50. The Cf-252 was now clearly visible. This was a stark contrast to the result obtained without PSD in Figure 5.48.

5.7.2 Neutron Detection Performance

With the data and postprocessing tools described in the previous sections, it was finally possible to report on the performance of the DDLI PoC. One of the major goals of the DDLI project was to understand the benefit added by neutron detectors in a mobile threat detection platform. While the DDLI was capable of many detection modalities, the results presented in this section will be specifically focused on neutron counting.

Under certain assumptions, it is possible to calculate the optimal integration window for a counting measurement when a detector is driven past a source. A treatment of this problem can be found in [47], but the results will be summarized here. In the case of constant background, the optimal integration is over a 108° cone

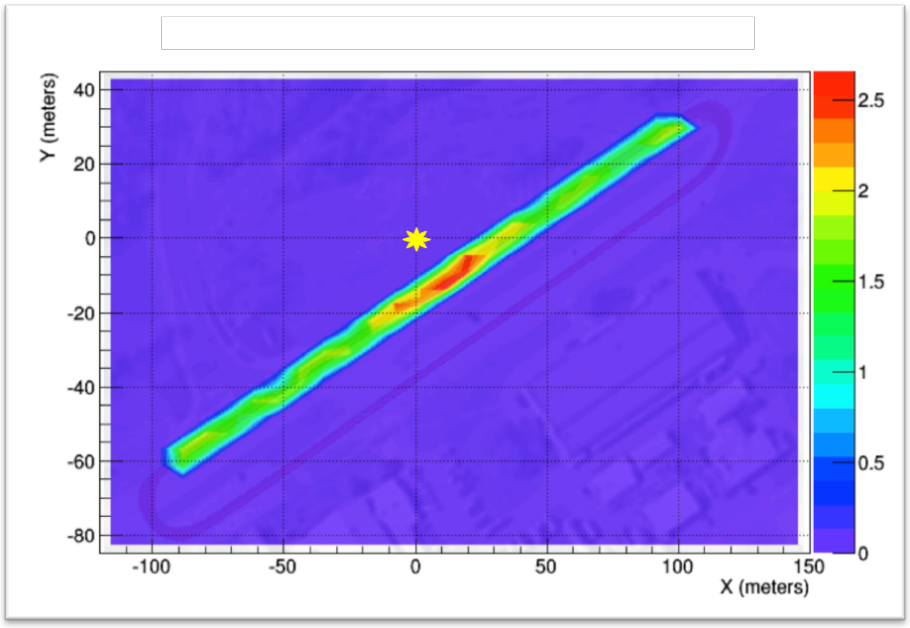


Figure 5.50: Neutron count rate map for the Cf-252 measurement. This map was generated using the organic scintillator events above 200 keVee which were neutrons with at least 5σ confidence. The X and Y axes represent distance from the source in meters. The neutron count rate is represented by the colorbar in counts per second. The source is visible after rejecting gamma rays with PSD.

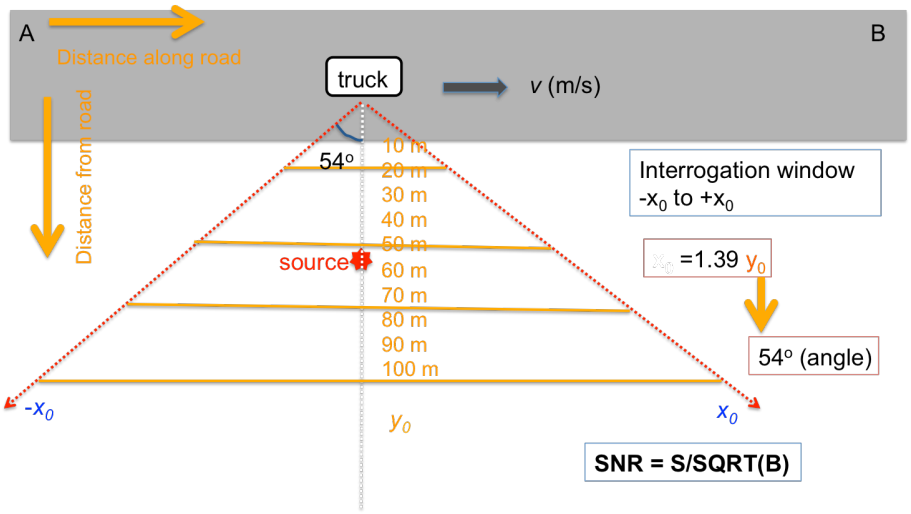


Figure 5.51: A diagram of of the optimal integration angle for counting. Diagram by B. Ayaz-Maierhafer.

centered on the source as depicted in Figure 5.51. The optimal integration length is not a function of speed. Rather, it is dependent on the standoff distance to the source.

Count rate maps were generated in Section 5.7.1.4 for both Cf-252 neutrons and the natural neutron background. Since these are maps of count rate, an integrated count can be calculated for a pass by the source at any speed. In the following analysis, the optimal window length will be used to report on the detection performance for the DDLI PoC at a variety of speeds. To clarify the conditions under which these results were obtained, an overview of the measurement will be presented.

On July 16, 2015, the DDLI PoC trailer was driven past a Cf-252 source placed in a field to the north of Bethel Valley Road at Oak Ridge National Laboratory. According to a measurement made with the Californium shuffler on September 11, 2015, the Cf-252 source emitted approximately 279,000 neutrons per second into 4π. At closest approach, the DDLI PoC trailer passed 15.9 meters from the source. This corresponds to an optimal counting window length of 43.8 meters. Neutrons were

Table 5.1: Estimated neutron count rates and their 95% confidence interval as a percentage.

	count rate (Hz)	% uncertainty (95%)
S+BG	2.74	14.6%
BG	1.59	6.4%

selected by taking a 5 sigma cut below the gamma ray band. Additionally, an energy window of 100 keVee to 1000 keVee was used. Background data was collected on July 7 and 8, 2015 at the same location.

To report on expected detection performance, the mean source-plus-background (S+BG) and background-only (BG) neutron count rates were estimated from measurements. To maximize the signal-to-noise ratio, the S+BG count rate was was estimated using data from the optimal integration window of 43.8 meters. The total neutron count in this window was low, however, leading to high uncertainty in the S+BG mean count rate. The background-only count rate was estimated somewhat differently. In this case, the entire background measurement dataset was used. Due to the lack of obstructions like large buildings, the neutron background count rate was expected to be constant throughout the Bethel Valley measurement site. By utilizing the all background measurement data, uncertainty in the mean BG count rate was greatly reduced.

The measured mean neutron count rates are listed in Table 5.1 along with their percent uncertainties. A $\pm 1.97\sigma$ interval on the count rates was used to determine a 95% confidence interval for the mean. The S+BG count rate was the largest contributor to uncertainty in the analysis of system performance.

With estimates for the mean S+BG and BG neutron count rates in hand, it was possible to estimate the number of counts the system would detect for a given speed. This quantity was calculated according to the following equation:

$$c = \frac{dr}{s} \tag{5.1}$$

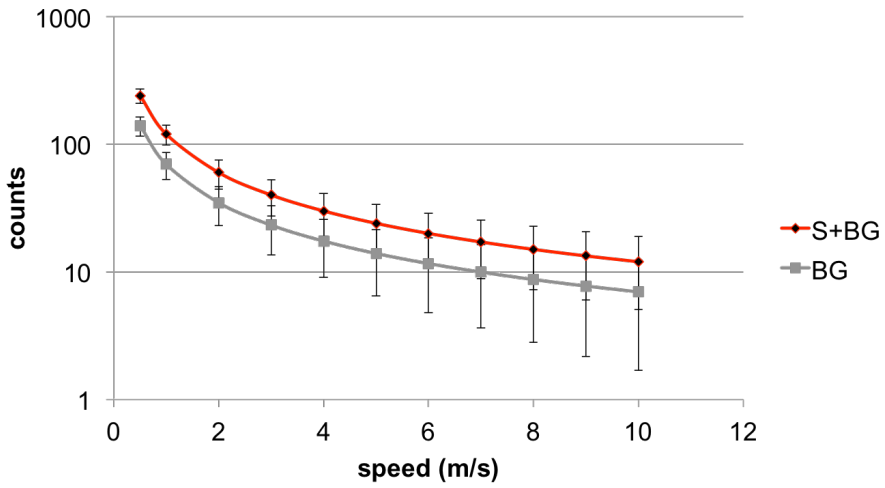


Figure 5.52: Expected neutron counts in the built PoC trailer for a variety of speeds.

where c is the number of counts, d is the integration distance, r is the mean count rate, and s is the speed.

Figure 5.52 shows the expected number of neutron counts as a function of speed for the Cf-252 source measurement and background. The error bars on this plot cover a range of $\pm 2\sigma$ as determined by Poisson statistics and the measured count rate. Even with the partially instrumented PoC trailer, fairly good separation between the source and background populations is achievable at slow driving speeds.

A fully instrumented system would be expected to provide better performance. The built PoC has only 6% of the plastic scintillators called for in the fully instrumented, double-sided PoC design. In terms of Liquid scintillators, only 23% of the central array is populated. In order to understand the detection performance of the scaled-up PoC design, measured count rates were scaled to reflect this increased detector area. This scaling was done on a per-detector-type basis. To account for self shielding, only the signal from one active mask was considered.

Figure 5.53 shows the expected number of counts in the scaled system as a function of speed. As can be seen, the $\pm 2\sigma$ error bars are significantly separated, even at reasonable speeds. Figure 5.54 quantitatively characterizes this separation via the

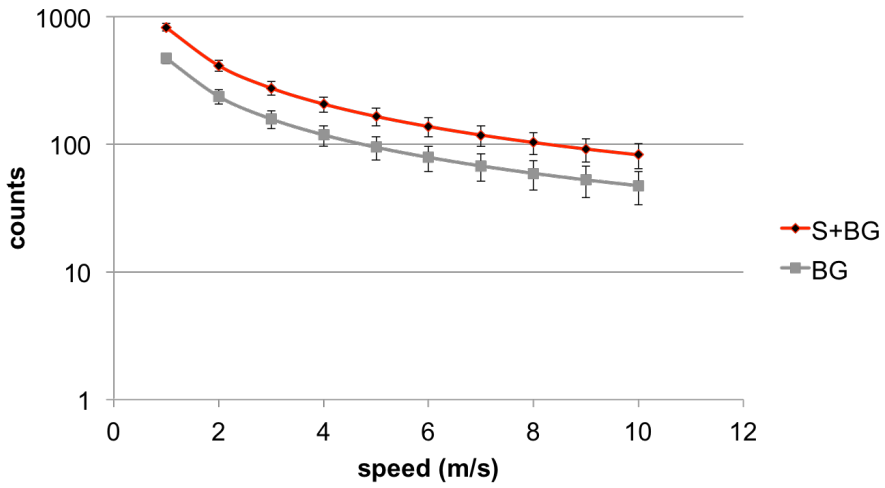


Figure 5.53: Expected neutron counts in the scaled-up PoC design for a variety of speeds. Error bars are $\pm 2\sigma$ and based on Poisson statistics from the scaled, measured count rate.

following figure of merit:

$$FOM = \frac{\mu_2 - \mu_1}{\sigma_1 + \sigma_2}. \tag{5.2}$$

Using this FOM, a larger number means more separation between the source and background populations. The error bars of Figure 5.54 represent 95% confidence intervals. These intervals were calculated by propagating 95% confidence bounds on the mean count rates (see Table 5.1).

Clearly the scaled-up system performs significantly better than the partially instrumented system, but what doe these FOM values mean in terms of detection performance? Figure 5.55 shows the detection rate vs speed for a number of different thresholds above background in the scaled-up PoC. Poisson statistics were used to calculate the probability of detection given a threshold. Poisson statistics can also be used to determine the false positive rate induced by the chosen threshold. According to this analysis, detection rates of 99.3% are expected with a 4σ threshold above background in a fully instrumented system when driving 5 meters per second. Figure

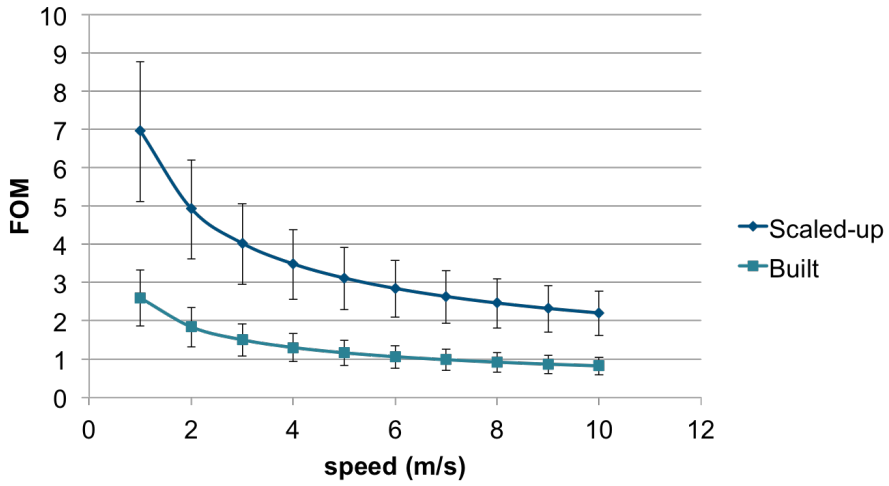


Figure 5.54: Comparison of detection performance in the built and scaled-up PoC designs for a range of speeds. The y-axis shows a figure of merit describing the separation of source and background populations. Error bars represent 95% confidence bounds.

5.56 shows the 95% confidence boundaries for the 4σ alarm threshold. At 5 m/s, the lower bound is still quite high at 81.5%. More detection rates for the built and scaled-up systems are listed in Tables 5.2 and 5.3.

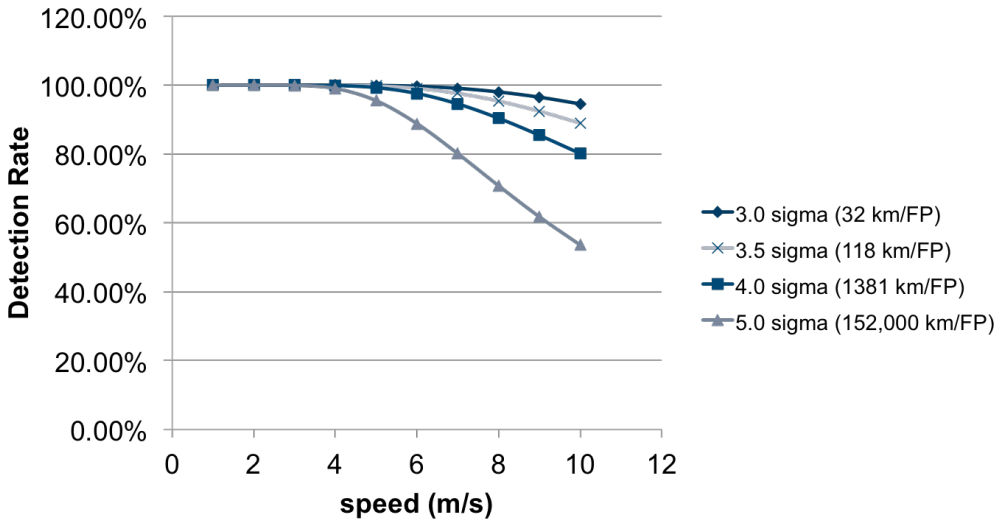


Figure 5.55: Detection probability as a function of speed in the scaled-up PoC when using neutron counting alone. Results are shown for a number of alarm thresholds. For values and confidence intervals, see Table 5.3.

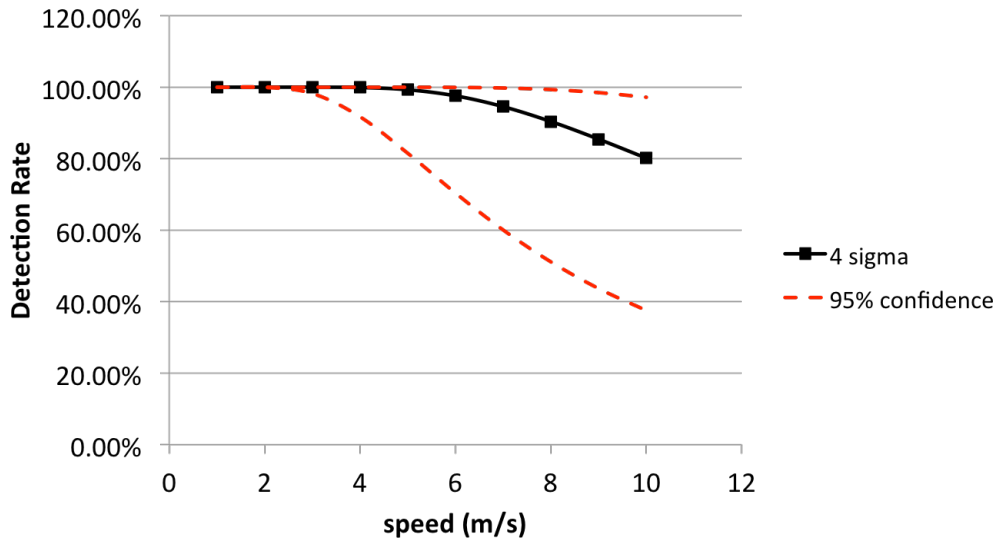


Figure 5.56: Detection probability and 95% confidence bounds as a function of speed for a 4σ neutron counting alarm in the scaled-up PoC.

Table 5.2: Detection performance as a function of speed for the built PoC system. Results are shown for 4σ and 5σ alarm thresholds. The low and high bounds represent the limits of 95% confidence interval.

	4σ			5σ		
speed (m/s)	Low Bound	Expected	High Bound	Low Bound	Expected	High Bound
1	54.6%	93.9%	99.8%	24.1%	78.3%	98.3%
2	19.0	57.9	88.0	4.5	28.7	67.9
3	9.3	34.6	67.0	1.6	12.3	39.3
4	5.7	22.6	50.0	0.8	6.5	23.8
5	3.9	16.0	38.2	0.5	4.0	15.6
6	2.9	12.0	30.1	0.3	2.6	10.9
7	2.3	9.5	24.4	0.2	1.9	8.0
8	1.9	7.7	20.2	0.2	1.4	6.1
9	1.6	6.5	17.2	0.1	1.1	4.9
10	1.4	5.5	14.8	0.1	0.9	4.0

Table 5.3: Detection performance as a function of speed for the scaled-up PoC system. Results are shown for 4σ and 5σ alarm thresholds. The low and high bounds represent the limits of 95% confidence interval.

	4σ			5σ		
speed (m/s)	Low Bound	Expected	High Bound	Low Bound	Expected	High Bound
1	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
2	100.0	100.0	100.0	99.4	100.0	100.0
3	98.2	100.0	100.0	90.2	99.9	100.0
4	91.7	99.9	100.0	71.7	99.0	100.0
5	81.5	99.3	100.0	53.4	95.5	99.9
6	70.4	97.6	99.9	39.1	88.8	99.5
7	60.1	94.6	99.8	28.9	80.1	98.3
8	51.1	90.4	99.3	21.6	70.8	96.1
9	43.7	85.5	98.5	16.6	61.8	92.7
10	37.5	80.1	97.2	12.9	53.6	88.5

5.8 Conclusions

This chapter has described the design, build, and testing of the DDLI, a mobile neutron and gamma ray coded aperture system. The main focus of this work has been to provide detection performance results for a proof of concept DDLI prototype. Specifically, the DDLI PoC was built to examine the benefits of incorporating neutron detection capability into a mobile threat detection platform.

The DDLI PoC was constructed over the course of a year at Oak Ridge National Laboratory. The build process involved extensive work with radiation detectors, data acquisition, digital signal processing, GPS systems, and software design. After the completion of the build, a measurement campaign was conducted to characterize system performance.

Analysis of the measured data revealed that good detection performance could be obtained in a fully instrumented PoC system using neutron counting alone. For a 3×10^5 neutrons/second Cf-252 source in constant background, detection rates of 99.3% were expected at 10 m/s and 15.9 meter standoff. These results were consistent with earlier simulations indicating the strength of neutron counting when the background is well behaved.

PSD efficiency in the DDLI PoC was relatively poor (see Appendix D). This was the result of problems specific to the system as well limitations in the existing PSD scintillator materials. Improved results should be expected if PSD performance could be increased, particularly at low energies. This is due to the significant response expected from Cf-252 at low levels of energy deposition.

The spontaneous fission neutron spectrum in Cf-252 is well modelled by a Watt distribution [48]. The Watt spectrum is given as

$$\chi(E) = \frac{e^{-(E+E_w)/T_w}}{\sqrt{\pi E_w T_w}} \sinh \frac{4E_w E}{T_w^2} \tag{5.3}$$

where E is the neutron energy in MeV, $\chi(E)$ is the fraction of neutrons emitted per MeV per fission, and E_w and T_w are empirically determined fit parameters [49]. For Cf-252, E_w and T_w are 1.175 and 0.359 respectively [48]. According to this model, approximately 30% of fission neutrons are emitted below 1 MeV. In EJ-299 plastic scintillator, a recoil proton at 1 MeV yields of approximately 110 keVee of light [50]. This was well below the effective range for PSD in the DDLI PoC.

This loss of signal is compounded by the response functions of the organic scintillators. Most neutrons will not deposit their full energy within the detectors. Instead, energy deposition will fall within a continuum between zero and the full neutron energy [51]. Therefore, even neutrons with incident energies above the effective PSD threshold cannot be identified some fraction of the time.

For these reasons, future work with the DDLI PoC should target an improvement in PSD capability at low energies. This could be achieved by reducing electronic noise, improving fitting functions, or rejecting pileup events. It may also be possible to realize performance improvements through machine learning PSD. Some work has already been published to this end [52]. In future systems, new PSD capable detector materials may help push detection to lower energy regions.

Chapter 6

Conclusions

This document has described the design, build, and testing of a mobile, neutron/gamma-ray, coded aperture imaging system. Over the past year, a proof of concept version of the Dual-Detection-Localization-Identification (DDLI) system was completed Oak Ridge National Lab. A series of mobile measurements were conducted with this device to investigate the utility of neutron sensing in a mobile platform.

These measurements indicated that high detection rates would be possible using neutron counting alone in a fully instrumented system. For a 280,000 neutrons per second Cf-252 source placed 15.9 meters away, a 4σ detection rate of 99.3% was expected at 5 m/s (over 11 miles per hour). These results support the conclusion that neutron sensing enhances the detection capabilities of systems like the DDLI when compared to gamma-only platforms. For threat classes with weak or heavily shielded gamma ray signatures, neutron sensors may be the only means of detection.

In addition to the PoC build, advanced algorithm research was a major topic of this work. To this end, machine learning algorithms were developed to improve the baseline detection capabilities of mobile imaging systems. The fusion of neutron and gamma ray coded aperture images was shown to outperform either modality

alone through a simulated study. In a simulated 1-D coded aperture imaging dataset, machine learning algorithms using both neutron and gamma ray data outperformed gamma-only threshold methods for alarming on weapons grade plutonium.

In a separate study, a Random Forest classifier was trained on a source injection dataset from the Large Area Imager, a mobile gamma ray coded aperture system. Geant4 simulations of weapons-grade plutonium (WGPu) were combined with background data measured by the Large Area Imager to create nearly 4000 coded aperture images. At 30 meter standoff and 10 m/s, the Random Forest classifier was able to detect WGPu with error rates as low as 0.65% without spectroscopic information. A background subtracting filter further reduced this error rate to 0.2%. Finally, a background subtraction method based on principal component analysis was shown to improve detection by over 150% in a figure of merit.

In general, the demonstrated results in neutron sensing and machine learning show that significant room for improvement exists in mobile threat detection. This improvement can come in the form of new sensing modalities or new data analysis techniques. In either case, making use of more information within the data stream or environment is the underlying cause of any gains. Although the pathways taken within this work were mostly oriented toward proof of concept, they were still able to produce good results. In light of this, "low hanging fruit" is almost certainly available for any researchers willing to optimize and dig deeper.

In the future, the work of this dissertation could be improved in many ways. As an example, the machine learning results would benefit from randomized feature selection. Given the set of features that were used, it would be of great benefit to demonstrate a general trend of improved detection performance with increasing feature set size.

Background subtraction in coded aperture images could also be revisited. An excellent exercise would be to develop the optimal frequency filter and compare its results to the filter from this work. It also remains an open question how PCA background subtraction affects the performance of the Random Forest classifier.

Along the same vein, it is likely possible to use a more effective subset of PCA components in the subtraction process. There is no reason to believe that eigenvalue ordering is optimal when trying to preserve source signal.

In terms of the neutron counting results, some of the potential work is obvious. In the reported measurements, count rates were simply too low to provide useful confidence intervals for detection beyond a few meters per second. To understand the capabilities of the system at higher speeds, longer measurements or more intense sources will be required. To truly know how the system will perform in the real world, however, validated source injection studies should be conducted. In addition to quantifying system performance, studies of this nature would help shed more conclusive light on the relative benefits of neutron counting versus neutron coded aperture.

Finally, it is important to tie together the machine learning and neutron sensing avenues of research. The DDLI system was designed as a hybrid, gamma-ray and neutron imaging platform. How then does the inclusion of such a demonstrably strong detection pathway like neutron sensing in a gamma ray system improve overall performance? It has been shown in this dissertation through simulation that significant performance gains should be expected from fusing these two sources of information. Still, this effect remains to be quantified in measurement.

At the conclusion of the project, the DDLI PoC system will be transferred to the University of Tennessee. There, the detector platform will be available as a tool for possible later research. Additionally, it has the potential to be useful as an instrument in student labs with the Nuclear Engineering department. Hopefully, access to the DDLI system will provide insight for students interested in nuclear security research.

Bibliography

[1] KP Ziock, JW Collins, L Fabris, S Gallagher, BKP Horn, RC Lanza, and NW Madden. Source-search sensitivity of a large-area, coded-aperture, gamma-ray imager. *Nuclear Science, IEEE Transactions on*, 53(3):1614–1621, 2006. xv, 8, 70

[2] Frezghi Habte, Mark F Cunningham, Lorenzo Fabris, and Klaus P Ziock. Performance of a large-area, gamma imager using a point source injection technique. In *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*, volume 2, pages 1157–1161. IEEE, 2007. xvii, 70, 71, 72

[3] RD Penny, WE Hood, RM Polichar, FH Cardone, LG Chavez, SG Grubbs, BP Huntley, RA Kuharski, RT Shyffer, Lorenzo Fabris, et al. A dual-sided coded-aperture radiation detection system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 652(1):578–581, 2011. xvii, 73, 74, 75

[4] Andrew Hoover, Mark Wallace, Mark Galassi, Michal Mocko, David Palmer, Larry Schultz, and Shawn Tornga. Simulation and modeling for the stand-off radiation detection system (sords) using geant4. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pages 914–917. IEEE, 2009. xvii, 76

[5] Lee J Mitchell, BF Phlips, W Neil Johnson, Eric A Wulf, Anthony L Hutcheson, CJ Lister, Kelia D Bynum, Byron E Leas, and Gerald Guadagno. Mobile imaging

and spectroscopic threat identification (misti): system overview. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pages 110–118. IEEE, 2009. xvii, 77, 78

[6] Timothy J Aucott, Mark S Bandstra, Victor Negut, Joseph C Curtis, Daniel H Chivers, and Kai Vetter. Effects of background on gamma-ray detection for mobile spectroscopy and imaging systems. *Nuclear Science, IEEE Transactions on*, 61(2):985–991, 2014. xvii, 17, 78, 79, 80, 81

[7] Nick Mascarenhas, Jim Brennan, Kevin Krenz, Jim Lund, Peter Marleau, Julia Rasmussen, Jim Ryan, and John Macri. Development of a neutron scatter camera for fission neutrons. In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 1, pages 185–188. IEEE, 2006. xviii, 79, 82

[8] Peter Marleau, James Brennan, Kevin Krenz, Nicholas Mascarenhas, and Stanley Mrowka. Advances in imaging fission neutrons with a neutron scatter camera. In *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*, volume 1, pages 170–172. IEEE, 2007. xviii, 79, 82

[9] Erik Brubaker. Fast neutron detection and imaging for nuclear security applications. (November), 2011. xviii, 79, 83

[10] Paul Hausladen and MA Blackston. Passive and active fast-neutron imaging in support of afci safeguards campaign. Technical report, Oak Ridge National Laboratory Report ORNL/TM-2009/210, 2009. xviii, 83, 84, 85

[11] A Iyengar, M Beach, RJ Newby, L Fabris, LH Heilbronn, and JP Hayward. Systematic measurement of fast neutron background fluctuations in an urban area using a mobile detection system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 773:27–32, 2015. xix, 91, 93

[12] DL Upp and RM Keyser. Performance of a car-mounted neutron and gamma-ray monitoring system for illicit material detection. In *INMM 45th Annu. Meeting*, 2004. 9

[13] El E Fenimore and TM Cannon. Coded aperture imaging with uniformly redundant arrays. *Applied optics*, 17(3):337–347, 1978. 10

[14] Prateek Tandon, Peter Huggins, Artur Dubrawski, Simon Labov, and Karl Nelson. Detection of radioactive sources using bayesian aggregation of data from mobile spectrometers. 14, 15, 32

[15] Simon Labov. Enhanced threat detection using machine learning data fusion analysis. In *Nuclear Science Symposium, 2012. IEEE*. IEEE Symposium on Radiation Measurements and Applications, 2012. 16

[16] Simon Labov. False alarm suppression of radiation portal monitor measurements using machine learning analysis of spatial signatures. In *Nuclear Science Symposium, 2012. IEEE*. IEEE Nuclear Science Symposium, 2012. 16

[17] Tracy D Lemmond, Barry Y Chen, Andrew O Hatch, and William G Hanley. An extended study of the discriminant random forest. In *Data Mining*, pages 123–146. Springer, 2010. 17

[18] David Boardman, Mark Reinhard, and Alison Flynn. Principal component analysis of gamma-ray spectra for radiation portal monitors. *Nuclear Science, IEEE Transactions on*, 59(1):154–160, 2012. 18, 32

[19] David Boardman and Alison Flynn. A gamma-ray identification algorithm based on fisher linear discriminant analysis. *Nuclear Science, IEEE Transactions on*, 60(1):270–277, 2013. 19, 32

[20] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification. 2nd. *Edition. New York*, 2001. 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 32

[21] Thomas M Cover and Peter E Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967. 21

[22] David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991. 21

[23] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is nearest neighbor meaningful? In *Database TheoryICDT99*, pages 217–235. Springer, 1999. 23

[24] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984. 25

[25] Vladimir Naumovich Vapnik and Samuel Kotz. *Estimation of dependences based on empirical data*, volume 40. Springer-verlag New York, 1982. 26

[26] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 26

[27] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593–605. IEEE, 1989. 28

[28] Nathan Intrator. Feature extraction using an unsupervised neural network. *Neural Computation*, 4(1):98–107, 1992. 28

[29] Geoffrey E Hinton, Peter Dayan, Brendan J Frey, and Radford M Neal. The" wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214): 1158–1161, 1995. 28

[30] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006. 28

[31] Harry Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004. 29

[32] Anand Narasimhamurthy. Theoretical bounds of majority voting performance for a binary classification problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1988–1995, 2005. 29

[33] YS Huang and CY Suen. The behavior-knowledge space method for combination of multiple classifiers. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 347–347. INSTITUTE OF ELECTRICAL ENGINEERS INC (IEEE), 1993. 29

[34] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006. 30, 49

[35] Yi Lin and Yongho Jeon. Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, 101(474):578–590, 2006. 30

[36] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. 30

[37] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1):37–52, 1987. 32

[38] Max Welling. Fisher linear discriminant analysis. *Department of Computer Science, University of Toronto*, 3, 2005. 32

[39] Snežana Dragović and Antonije Onjia. Classification of soil samples according to their geographic origin using gamma-ray spectrometry and principal component analysis. *Journal of environmental radioactivity*, 89(2):150–158, 2006. 32

[40] Israel Dov Cohen and Wondwosen Mengesha. Anomaly metrics to differentiate threat sources from benign sources in primary vehicle screening. Technical report, Sandia National Laboratories, 2011. 32

[41] MS Gordon, P Goldhagen, KP Rodbell, TH Zabel, HHK Tang, JM Clem, and P Bailey. Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground. *Nuclear Science, IEEE Transactions on*, 51(6):3427–3434, 2004. 35, 91

[42] Michael V Hynes, Maurice Toolin, Bernard Harris, John McElroy, Mark S Wallace, Larry J Schultz, Mark Galassi, Andrew Hoover, Michal Mocko, David Palmer, et al. The Raytheon-SORDS trimodal imager. In *SPIE Defense, Security, and Sensing*, pages 731003–731003. International Society for Optics and Photonics, 2009. 76, 77

[43] Lee J Mitchell, Bernard F Phlips, Eric A Wulf, Anthony L Hutcheson, and Byron E Leas. Cross country background measurements with high purity germanium. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pages 319–323. IEEE, 2011. 78

[44] Rene Brun and Fons Rademakers. Rootan object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1):81–86, 1997. 104

[45] RJ Livesay, Christopher S Blessinger, TF Guzzardo, and PA Hausladen. Rain-induced increase in background radiation detected by radiation portal monitors. *Journal of environmental radioactivity*, 137:137–141, 2014. 130, 131

[46] Marvin L Roush, MA Wilson, and William F Hornyak. Pulse shape discrimination. *Nuclear Instruments and Methods*, 31(1):112–124, 1964. 138

[47] Timothy J Aucott, Daniel H Chivers, and Kai Vetter. Proximity localization with the mobile imaging and spectroscopic threat identification (misti) system. In *2011 IEEE Nuclear Science Symposium Conference Record*, 2011. 141

[48] FH Fröhner. Evaluation of 252 cf prompt fission neutron data from 0 to 20 mev by watt spectrum fit. *Nuclear Science and Engineering*, 106(3):345–352, 1990. 150, 151

[49] J Kenneth Shultis and Richard E Faw. Radiation shielding. american nuclear society. *Inc., La Grand Park, Illinois USA*, 2000. 151

[50] S Nyibule, E Henry, WU Schröder, J Tõke, L Acosta, L Auditore, G Cardella, E De Filippo, L Francalanza, S Gìani, et al. Radioluminescent characteristics of the ej 299-33 plastic scintillator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 728:36–39, 2013. 151

[51] Kazuo Shin, Yoshitomo Uwamino, and Tomonori Hyodo. Propagation of errors from response functions to unfolded spectrum. *Nuclear Technology*, 53(1):78–85, 1981. 151

[52] Emanuele Ronchi, P-A Söderström, Johan Nyberg, E Andersson Sundén, Sean Conroy, Göran Ericsson, Carl Hellesen, M Gatu Johnson, and Matthias Weiszflog. An artificial neural network based neutron–gamma discrimination and pile-up rejection framework for the bc-501 liquid scintillation detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 610(2):534–539, 2009. 151

[53] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, et al. Geant4a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. 165

[54] K. Nelson and P. Sokkappa. A Statistical Model for Generating a Population of Unclassified Radiation Signatures Spanning Nuclear Threats. *LLNL-TR-408407*, 2008. 173

[55] Marco Cecconello, M Donato, C Marini-Bettolo, Sean Conroy, Siriyaporn Sangaroon, and Göran Ericsson. Measurement of the γ-ray energy resolution function of ej301 liquid scintillator using a dual channel adc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 753:34–37, 2014. 181

Appendix

Appendix A

Software Development

A.1 Gex

A.1.1 Geant4

Geant4 is a powerful Monte Carlo general particle transport tool written in C++ [53]. It was developed as an international collaborative effort and has been used in numerous projects across the world. Geant4 is fairly unique due to the fact that simulations are written by interacting with the Geant4 library in C++ code. Because Geant4 adopts an "object oriented" design, it is possible to exert great control over how a simulation proceeds and how data is extracted and managed. It supports highly customizable geometry, physics, source and material models. Because of these qualities, Geant4 was chosen as the simulation package for this project.

A.1.2 Previous simulation workspace

A simulation code had previously been developed for the early stages of this project in Geant4. This code was used to simulate the optimized DDLI design for source as well as background runs. A dynamic background was modeled using a constant background spectrum with varying intensity. This intensity was sampled at set

intervals from a distribution obtained during the Large Area Imager study, which involved driving the imager cross-country (see Section 4.6.2). The data extracted from these simulations was saved to disk using the serialization capabilities of CERN's ROOT data analysis package. 1D coded aperture images at a set depth were then reconstructed offline using a special ROOT script.

While this code was functional, it was not without its problems. Developed incrementally over the course of this and other projects, the design was somewhat haphazard. Because of tight coupling between components, adding new functionality was difficult and disabling old functionality was practically impossible. The existing scheme for saving simulation data was deeply intertwined with the various simulation components and the coded aperture reconstruction algorithms. This meant that modifying one component could easily have important and potentially unseen repercussions on the execution of other components.

Ignoring the software engineering viewpoint, one very practical problem with the old Geant4 workspace was that it was slow. Too much data flow and deep branching caused simulations to execute sluggishly, especially in the case of the gamma ray background which could take days to perform a single pass. This was more than a convenience issue. Using the existing data simulated with this workspace, machine learning had already shown promise as a way to accomplish the project's data fusion goals. However, machine learning techniques require large amounts of data for both training and validation. It would simply be infeasible to generate tens or hundreds of thousands of source passes using the existing Geant4 workspace due to the time scale required, even with much more computing power at our disposal.

A.1.3 New workspace

The new Geant4 version 10.0 introduced multithreading capabilities which were long overdue to the Monte Carlo code. In order to take advantage of these new capabilities, large changes needed to be made to our simulation workspace. These changes were deemed to be worth the effort for a number of reasons.

The first obvious reason was to gain some of the speedups promised by multithreaded execution. Given the large amount of simulated data needed to train machine learning alarming algorithms, increases in simulation speed are highly desirable. In addition to the speedup gained by distributing the simulation across CPU cores, the advantage of running a multithreaded instance of Geant4 is that the large datasets used in radiation transport calculations need not be loaded for each thread. This is not the case when multiple instances of the Geant4 process are executed simultaneously to achieve the same effect. This reduction in memory usage can mean better cache performance.

The second reason the changes were deemed necessary was to allow the workspace to be reorganized. Reorganization was necessary to make extending, debugging, and linking the simulation code with an analysis chain less difficult. To this end, it was decided to include many of the new features introduced in C++11 such as smart pointers to handle memory management and standard library mutexes to protect critical sections.

The new framework was designed with extendability and modularity in mind. However, Geant4's system of user hooks is somewhat resistant to writing code in this fashion. To combat this problem, a series of adaptors was created as the basis of the framework to expand the single-slot nature of Geant4's user hooks. Thus, any number of user actions, sensitive detectors, primary generators, and user informations can be used at a time. This functionality helps allow discrete tasks to be separated

into equally discrete pieces of code. A composite structure was also introduced for simulating sources which abstracts away different, typically-independent aspects such as angular, energy, and position distributions.

Data storage was also modified. Cern's ROOT package was still employed for I/O, but as it is not thread-safe, steps needed to be taken to ensure it would function correctly in the multithreaded simulation environment. To ensure the framework was as extendable as possible, a method for generating and linking ROOT dictionaries at runtime was employed in combination with type erasure. This method allows for any combination of data to selected for storage at runtime. This combination of type erasure and runtime dictionary compilation completely decouples the framework from the data and gives the user control over what they choose to save.

This framework was abstracted into an open source library called Gex, which is available for free on GitHub. The Gex library provides a set of useful classes in addition to the underlying framework. These tools allow a multithreaded Geant4 simulation, complete with data output, to be written in just a few lines of code. This is accomplished by simply including modular functionality, such as particle identification, into the simulation. Since the framework is extensible, advanced users can customize these tools or write their own as necessary.

A.1.4 Performance improvements

Using a Lubuntu virtual machine with 1GB of memory and the DDLI prototype geometry (Section 5.3.1), a WGPu source was simulated for a 30 meter standoff pass with both the new and old workspaces to compare run time. Geant4 10.0 was used in both cases. The old workspace required 87 minutes to simulate the pass, while the new workspace required only 4 minutes when using a single thread. This indicates that the new workspace is significantly more efficient even without using multithreading. When using 4 threads, the new workspace completed the simulation in just 1 minute 23 seconds, representing around a 60× speedup over

the old workspace. This translates to the ability to simulate 1000 gamma source passes in 23 hours with the new workspace vs. around 2 months with the old on this particular machine.

A.2 Pipe

In the course of this project, several different data streams needed to be combined. Large amounts of simulated data were injected into data taken by multiple real data acquisition systems. Data from these real systems needed to stream live or be recovered offline from files. Once the data had been loaded properly, it was processed in ways that are specific to the system, such as coded aperture imaging reconstruction, proximity localization imaging, counting, and spectroscopy. Additionally, streaming needed to continue indefinitely, as it is not known beforehand how long or far the systems will be in operation.

In order to support these requirements, a multithreaded framework for data analysis was created in C++11. The framework takes the form of a pipeline to process data from an input stream. This stream can be real-time acquisition or replay from a file. Each module in the pipeline spawns its own thread and may read or attach data to type-erased data bundles which are passed from module to module. Modules sleep when they are not processing data. The high level of abstraction and pipeline setup allows modules to be developed as separate units and deployed in the combinations deemed necessary. Module configuration may be accomplished via XML if desired. A large number of modules have already been developed and deployed successfully for the study described in Section 4.6.

Appendix B

Original Plan

At the outset of experimental work, a plan was developed which would tie the physical build effort and measurements back into the simulation workspace. In this formulation, the DDLI system measurements would serve as a validation tool for Geant4 simulations. After system simulations had been validated, the simulations, as opposed to the measured data, would be used to report on system performance. In this section, this original plan will be described. Knowledge of the the planned effort will inform the direction of the PoC work.

The physical construction effort for the DDLI PoC began after the algorithmic studies in Chapter 4 had been conducted. The development of these algorithms and the assessment of benefit had been largely based on source injection datasets assembled from Geant4 simulations and measured background. The largest source injection study, which employed data from the Large Area Imager, had several shortfalls with regard to answering the questions posed by the project.

The first and largest problem was that the LAI system sensed gamma rays only. That meant that no injections of neutron sources into measured neutron background were possible during the study. Two of the projects major goals were to demonstrate the benefit of neutron sensing and neutron/gamma data fusion: clearly this dataset was insufficient for those tasks. Additional issues arose from the fact

that the Large Area Imager was driving at variable highway speeds throughout its background collection measurement, necessitating a time-sampling manipulation of the background data. Finally, energy calibration measurements were not conducted during the trip. This forced the studies to focus on gamma-ray counting only, without spectroscopy.

Nevertheless, the LAI provided a good model for a source injection study. If a new dataset could be constructed which fixed the problems mentioned in the last paragraph it would allow for very good estimations of system performance to be obtained. The DDLI PoC would, of course, contain neutron detectors. If a large background dataset could be obtained using the PoC system, neutron source injection studies would be possible. The issues of speed and spectroscopy could also be solved through operational procedure and some data processing. With this in mind, the following plan of action was adopted:

1. Construct DDLI PoC system and bring it online
2. Conduct a series of validation measurements with the PoC
3. Recreate measurements in Geant4 and adjust model until there is good agreement
4. Collect a large body of background data with the PoC under controlled conditions
5. Use validated Geant4 model to simulate a population of potential threat sources
6. Create a source injection dataset from measured background and simulated threats
7. Train detection algorithms following the procedure used for the Large Area Imager study
8. Report on expected performance of a scaled-up system based on the PoC

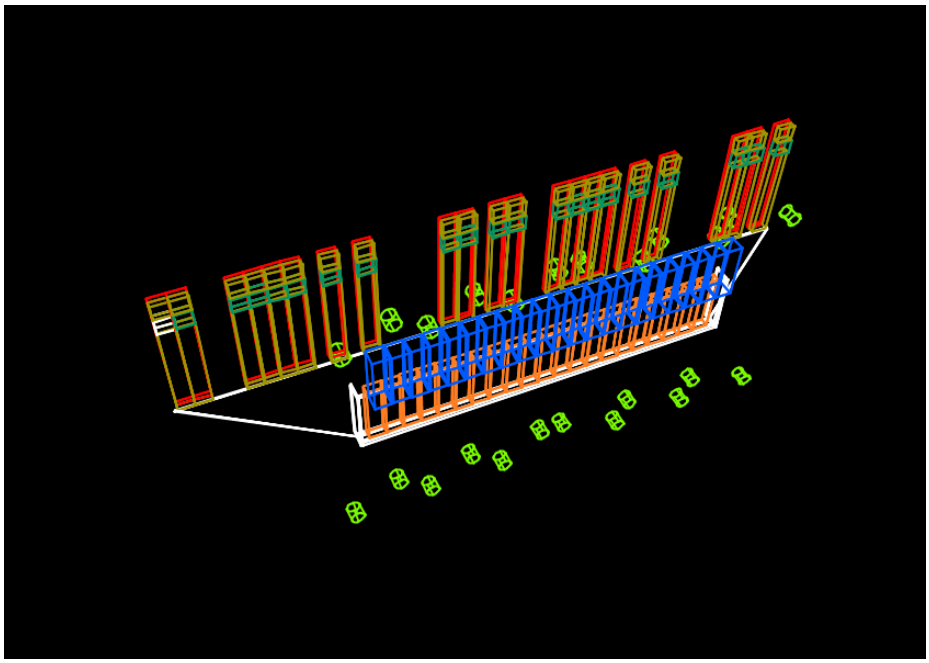


Figure B.1: Geant4 model of the DDLI PoC system.

B.1 Geant4 Model

To facilitate item (3) in the original plan, a model of the DDLI PoC was created in Geant4. This model was a modification of the existing model for the optimized DDLI system. It would be further reduced to match the partially instrumented DDLI PoC which was actually constructed. Measurements with the PoC would then be used to validate the model's response. The PoC model is shown in Figure B.1.

B.2 Background Measurements

Simulating natural radiation background is extremely difficult, especially for a moving system. As has been shown by studies in this document, the background intensity can vary significantly over the range of a few of meters. Additionally, the background spectrum also changes. The spectral changes cannot be ignored in any analysis

that wishes to assess performance using spectral information from the system. For this reason, it was deemed preferable to measure background data instead of using simulations.

To acquire this data, the PoC trailer would be driven around Tennessee. The target goal would be to collect around of 1000 miles of background data over the course of one or two weeks. To fix the issues that made working with the Large Area Imager background data difficult, regular energy calibrations would be performed. Also, an attempt would be made to keep the imager at a constant slow speed around 25 mph.

B.3 LSNM Dataset

By the time the build phase was ramping up, Lawrence Livermore National Laboratory (LLNL) had developed an approach for generating a population of nuclear threat signatures [54]. Instead of simulating a single type of source, such as unshielded WGPu, as was done for the LAI study, we wished to simulate the DDLI system response for a wide variety of difficult threats. A simulation of this nature would make it clear to what extent neutron sensing improved performance. It would also shed light on the lower limits of the system in realistic scenarios where the source would almost certainly be shielded. Finally, it would provide an opportunity to show that some sources, detectable with neither gamma ray or neutron sensing alone, were visible using data fusion.

The L-SNM population is a set of 1-dimensional threat source models consisting of an internal void, fissile material, and up to four layers of shielding. HEU of different compositions, ^{233}U, depleted urainium (DU), various Plutonium mixtures, and Neptunium are all possible components in the fissile core of the model. The models were randomly generated using a sensible value range for each parameter. Models were then subject to rejection based on their overall characteristics. Models which either emitted no radiation or had a does rate greater than 100 microrem per

hour at one meter were rejected. Additionally, models which were near critical were also rejected. The criterion for rejection in this case was a multiplication factor of 0.9 or greater.

There were over 8000 1D SNM models in the database made available to us for the project. These models were converted into neutron and gamma ray flux as a function of energy using the radiation transport tool GADRAS. In order to perform the final evaluation for the alarm algorithms developed in this project, these source specifications would be used to perform source injection studies using the Geant4 workspace.

A major concern was the simulation time this effort would require. Even with the speedup from the new Geant4 workspace (see Section A.1.3), simulating 100 passes for each source would take more than 3 years on a quad core machine. Even 10 passes would require around 150 days, not including post processing. Given the time remaining in the project, this simulation time was not feasible. Given the time constraints, two options existed for generating the evaluation data.

The first was to generate a much reduced subset of the population. The difficulty would be determining the criteria for the selection of this subset. The strength of the gamma and neutron flux would likely be useful, though not completely informative in this case. The quality of the subset would be improved by considering the source type and shielding in the selection. If the subset was reduced to a tenth of its size, it would be possible to simulate 100 passes for each source in around one month with 4 quad core machines.

A second alternative was to scale the flux up to a level where high statistics could be achieved in a single simulated run. In this approach, the resultant data would then be decimated during source injection. If the time required to acquire good statistics was much less than the time required to simulate the desired number of passes, time would be saved.

Appendix C

Experiments to Support Simulations

As presented in Appendix B, the original project plan saw the partially instrumented DDLI PoC trailer as a tool for simulation validation. For this reason, a significant amount of experimental work was done in support of the simulation effort. There were certain aspects of the detector response which needed to be incorporated in an accurate simulation. The following section will discuss two experimental setups. The first was designed to give us an estimate of PSD efficiency as a function of energy. The second was designed to help us understand energy calibration and resolution in organic scintillators.

C.1 Time Tagged Cf-252

In our simulations, certain aspects of the detector system were not modeled. Some missing pieces of the puzzle were the generation of light in the scintillator; the collection of light by the photocathode; the propagation of the electrical signal through the digitizers; and the processing of the digital waveform. We did not simulate these aspects of an interaction because they are computationally intensive.

In the case of neutron detection, one high level property we wanted to simulate was our ability to discriminate neutrons from gamma rays. In real life, neutrons and gammas were differentiated using pulse shape discrimination (PSD). Because PSD

was performed on the digitized waveform, it was dependent on all of the missing components mentioned above. Of course, these components could be simulated all in one go by sampling from the probability distribution that governed the PSD outcome. If we knew the PSD outcome distribution, we could replace everything from the generation of light to the PSD processing with a single, computationally-cheap operation.

One way to learn the PSD outcome distribution is to use a Cf-252 fission chamber. A Cf-252 fission chamber is made a Californium source placed inside of a small gaseous detector. Cf-252 undergoes spontaneous fission. When a fission occurs, some number of neutrons and gamma rays will be emitted on average. Meanwhile, the chamber will emit a pulse due to energy deposition from the fission fragments. A time-of-flight TOF experiment can be constructed by using the pulse from the fission chamber as a "start" and the signal from a detector as a "stop." Because gamma rays travel at the speed of light, and neutrons do not, the TOF for an event provides a very good way to discriminate neutron events from gamma ray events. In a PSD detector, this "ground truth" can be used to evaluate the fraction of events which are correctly classified by PSD as a function of deposited energy. As a bonus, a neutron energy calibration can be developed in the same experiment. Since neutrons do not travel at the speed of light, the neutrons energy will determine its TOF.

The setup for the Cf-252 time of flight experiment required a few extra pieces of electronics. The fission chamber needed high voltage and a DC bias for the associated preamplifier. The signal from the fission chamber was then run into a constant fraction discriminator (CFD) to convert it to a NIM pulse. In our experiment, The NIM pulse and the signal from the detector being interrogated are fed into a SIS3316 digitizer. In order to capture the NIM pulse with good timing fidelity, it passed through a CR shaper box before reaching the digitizer. Without a CR shaper, the rising edge of the pulse would have fallen within a single 4 nanosecond sample, limiting the timing resolution to 4 ns. Since timing would be used to calculate neutron energy, a low timing resolution was desirable. Photos of the setup are shown in Figure C.1.

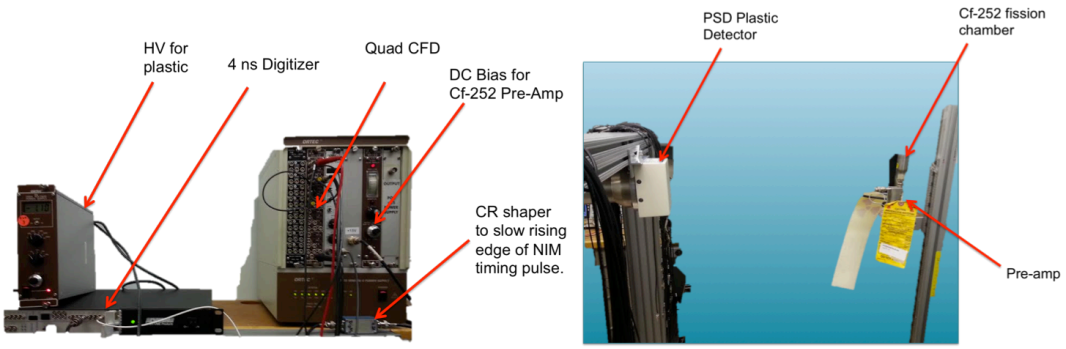


Figure C.1: Annotated photos of the electronics and detector setup for the time of flight measurements.

In addition to spontaneous fission, Cf-252 can also decay through alpha emission. Because alpha particles are have an electric charge like the fission fragments, they will also produce a pulse in the fission chamber. Fortunately, the pulse height distribution for alpha emissions is somewhat separated from the more energetic spontaneous fission events. Due to this fact, a pulse height threshold can eliminate most triggers from alpha emission. To determine this threshold for our experiment, a counting curve was made. To do this, the pulse height threshold on the CFD unit was incrementally increased. At each level, the number of NIM triggers generated per second was counted by a Scalar unit. A threshold value was then selected from the "plateau" region of the counting curve. The plateau indicates threshold values where most of the alpha distribution was suppressed. The generated counting curve is shown in Figure C.2

The resulting data from a TOF experiment with a plastic scintillator can be seen on the left side of Figure C.3. In this histogram, two populations are clearly visible. The vertical band near zero TOF contains the gamma ray events. Their time of flight is very low because they travel at the the speed of light. Neutrons arrive later, with the most energetic neutrons arriving soonest. The neutrons can deposit anywhere from no energy to their full energy in the plastic scintillator, leading to the wedge shape of the neutron population in this plot. The fastest neutrons (low TOF) can potentially deposit the most energy, leading to their high maximum light output in

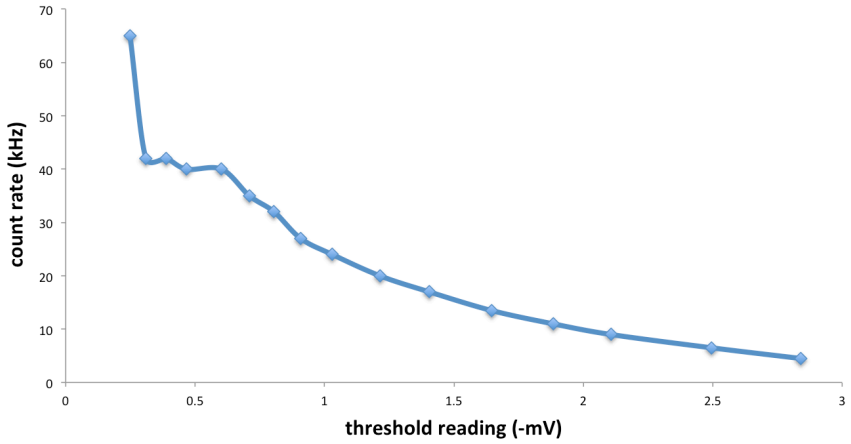


Figure C.2: The counting curve for the time-tagged Cf-252 chamber as a function of CFD threshold. A threshold from the plateau region was chosen to reject alpha events from the chamber.

the scintillator (MeVee). As the total neutron energy decreases (longer time of flight), the maximum possible deposited energy also decreases. In this way, the upper edge of the neutron population connects neutron energy (MeV) to light output (MeVee), and can be used as a neutron energy calibration.

If the same data is visualized with a PSD metric on the y-axis (right side of Figure C.3), a third population can be seen. These are accidental gamma rays caused by background or room scatter falling within the coincidence window. More importantly, this plot demonstrates why the time of flight experiment was important. In a normal measurement scenario, the TOF dimension is missing. Instead, we can only determine the PSD metric of a waveform and use it to make a guess about the true identity of the interacting particle. Without the TOF dimension, our data looks like a projection onto the PSD axis. As one can see, the neutron and gamma populations are overlapping in their possible PSD values. This means that there are certain PSD values for which we cannot be sure about the true identity of the particle – it could be a neutron or a gamma ray. In these cases, we typically discard the event, leading to reduced efficiency. The fraction of true neutrons rejected in this

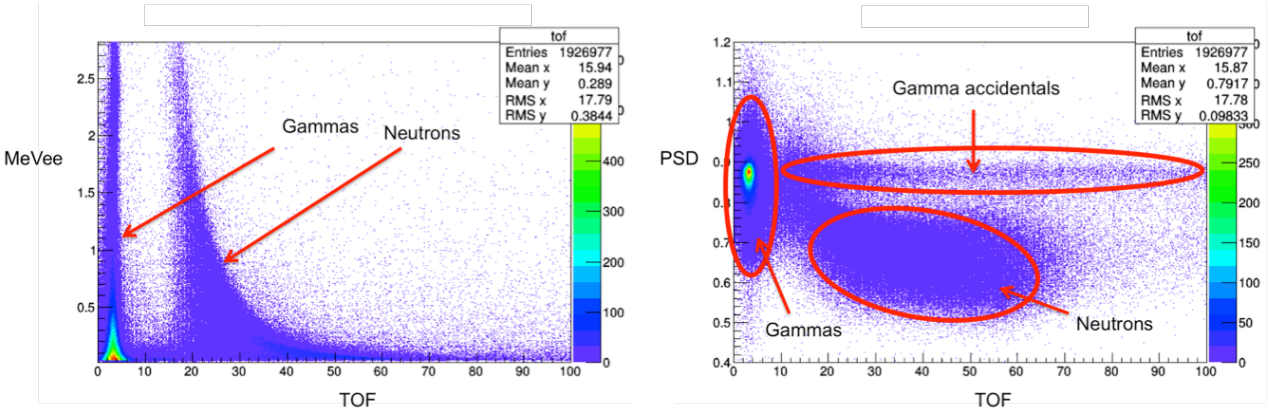


Figure C.3: The results of a time of flight measurement. On the left, a histogram of light output vs. time of flight shows the well separated gamma and neutron bands. On the right, a plot of PSD metric vs. time of flight reveals a third population of accidental gamma coincidences.

way determines the PSD efficiency. By discarding neutron events in simulation at a rate consistent with the PSD efficiency, it is possible to simulate PSD realistically without directly modeling scintillation, light transport, and electronics.

C.2 Compton Spectrometer

Determining an energy calibration in organic scintillators is difficult. This is because gamma rays rarely deposit their full energy in the low Z detector material. Instead, Compton scatter is more likely. In gamma spectra from materials such as NaI(Tl) of HPGe, full energy peaks are clearly visible. In this case, energy calibration is performed simply by finding the centroids of two or more full energy peaks and applying a fitting function. When working with organic scintillators, however, only the Compton continuum is available. In an ideal world, the Compton edge could be located just as simply as a full energy peak and used for calibration. Unfortunately, though, the Compton edge is obscured by several features which are difficult to control for.

Instead of a sharp cutoff at the Compton edge energy, a real spectrum will display a sloped edge due to multiple scattering interactions. The degree to which multiple

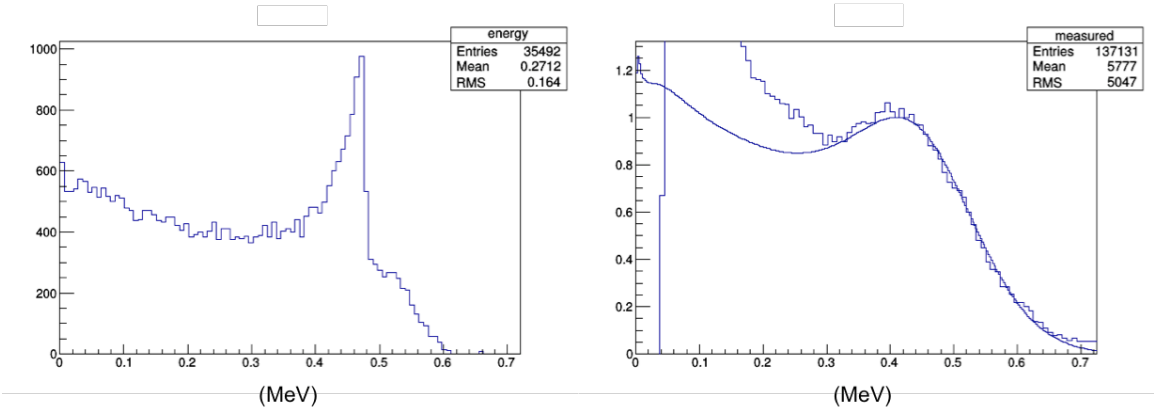


Figure C.4: (Left) a simulated Cs-137 spectrum in plastic scintillator without energy resolution effects. (Right) A measured Cs-137 spectrum in plastic (jagged line) and the left spectrum blurred to match (smooth line).

scatters are present is a function of the detector geometry and materials. The poor energy resolution of organic scintillators is another factor in distorting the edge. Figure C.4 shows a simulated Cs-137 energy spectrum in plastic without energy resolution effects. For comparison, Figure C.4 shows a measured energy spectrum from Cs-137 (jagged line). As you can see, the Compton edge has been highly smoothed into a continuous slope covering hundreds of channels.

When trying to generate an energy calibration from measured Compton edges, we are faced with a difficult question: which channel should be assigned the Compton edge energy? One way to get around the problem of picking a channel is to try to determine the energy calibration through simulation. If the detector, source, and surroundings are modeled to a sufficient degree of accuracy, it is possible to obtain a spectrum free of resolution effects, as has been demonstrated in Figure C.4. The units of this spectrum are in energy, not detector channels. If the simulated spectrum is blurred out according to the resolution of the detector, a spectrum matching measurements would be obtained with counts binned according to energy. Any spectral feature, such as the inflection point of the Compton slope, could then be mapped to both a channel number and an energy. Thus, an energy calibration is born.

When this procedure was attempted using our plastic scintillators, the initial results were promising. Figure C.4 shows a simulated Cs-137 spectrum (smooth line) matched to a measured Cs-137 spectrum (jagged line) in EJ-299. The simulated Compton edge matches the measured Compton edge quite well. There some disagreement at lower energies, but this is due to strong down-scatter from the room, which is discussed later. This result was obtained by blurring the simulated spectrum according to an empirically determined energy resolution function [55], the parameters of which were optimized according to an MSE fit to the measured Compton slope.

The problem with the fitted spectrum shown in C.4, is that an offset parameter was also introduced to obtain good agreement between the two spectra. Specifically, this parameter was designed to allow for the zeroth channel in the measured spectrum to not represent a nonzero energy. While this is commonly true of spectra generated using multichannel scalars (MCAs), our baseline subtracted waveform integrals were expected to have a true zero intercept. In plain language, we expected a the zeroth channel to really mean zero. Therefore, the fact that an offset was needed to get agreement between simulated and measured data was troubling.

One thing that could cause the need for an offset parameter would be a nonlinear relationship between energy and light output in our organic scintillator. Because the simulated spectrum is by definition linear in energy, any nonlinearity in the measured spectrum would cause the simulated Compton edge to require shifting to appear in the same location. Although this was investigated, organic scintillators are notoriously linear. This has been reported in literature several times.

Not being able to accurately simulate the response of a single detector sitting in a lab was cause for some concern. Eventually these detectors would be mounted in a complicated trailer. It would be much more difficult to troubleshoot a full system model. What was needed was a way to directly measure the energy calibration in our detectors. With some ground truth measurement, perhaps we could determine why

our simulations were not matching. Additionally, this ground truth measurement would be a much more rigorous way to determine the energy calibration for our organics.

To build an energy calibration, we need to map a known energy deposition to light output. Light output is measurable by integrating the detector pulse, but determining the deposited energy is more difficult, and is the reason why we seek a calibration in the first place. Compton scatter is the most likely gamma interaction in an organic scintillator as we have mentioned. Fortunately for us, it is possible to measure the deposited energy from a scatter interaction using a coincidence experiment.

If a gamma ray of known energy scatters in one detector and is stopped in a second, energy-calibrated detector, the energy deposited by the scatter may be calculated. It is simply the difference between the original gamma energy and the energy deposited in detector two. Gamma rays travel at the speed of light, so a short coincidence window should be able to separate out interactions that meet this condition. Because the energy deposited in a Compton scatter is a function of the scattering angle, an energy calibration can be constructed by moving the second detector to a different angular position. Of course, it is possible that the gamma ray will scatter in the second detector as well instead of depositing its full energy. This has the effect of creating a "full energy" peak for the scatter and a Compton continuum based on the scatter energy. This full energy band is most useful for calibration, because the width of the band is also related to the energy resolution of the scattering detector.

For our coincidence experiment, three liquid nitrogen cooled HPGe detectors were set up around an organic scintillator, as shown in Figure C.5. HPGe was chosen for its good energy resolution. The energy resolution of the stopping detector is important because it contributes to the uncertainty in the calculated energy deposition from scatter. The HPGe and organic scintillator signals were piped into a SIS3316 waveform digitizer. A Cs-137 source was placed near the organic scintillator to increase the chance of initial scatter in that detector.

Figure C.5: The setup of the Compton scatter experiment. Three HPGe detectors surrounded the organic detectors to be measured. The HPGe detectors were cooled by liquid nitrogen and powered by high voltage supplies with bias interlocks to prevent damage. Detector signals were recorded by a SIS3316 VME digitizer card.

The HPGe detectors used in this experiment are equipped with a preamplifier which integrates the signal from an event in the active volume. Because this signal is pre-integrated and contains a decay component, it must be processed differently than the organic scintillator to extract an energy spectrum. A post processing algorithm applying a trapezoidal filter was created for this purpose. The trapezoidal filter works by finding the difference between two moving window averages of the preamplifier signal. The goal of this operation is to find the baseline subtracted amplitude of the preamplifier pulse while averaging out noise before and after the rise. This pulse amplitude is proportional to the energy deposited in the detector.

Using this method, a plot such as the one in Figure C.6 can be created. This plot shows the HPGe energy vs the plastic scintillator energy (both in channels) for a Cs-137 source. The strong diagonal band represents scatter events in which the scattered gamma ray was fully stopped in the HPGe. The region below this band represents events where the scattered photon scattered again in the HPGe and escaped. Each

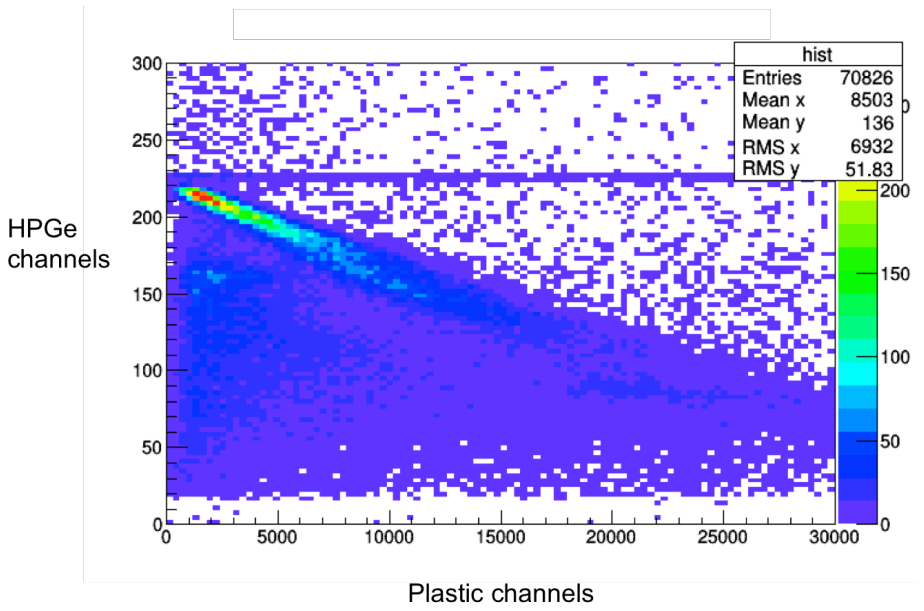


Figure C.6: A histogram showing the energy deposition in HPGe vs plastic for coincident events. The slope of the bright region determines the energy calibration while its width is related to the energy resolution in the plastic scintillator.

vertical slice in this graph would show a full energy peak at the scattered photon's energy and the associated Compton continuum. The thin horizontal band comes from accidental coincidences which deposited their full energy in the HPGe. It is a very thin band because of the excellent energy resolution of the HPGe detectors.

Using the diagonal full energy band and an energy calibration for the HPGe detectors, an energy calibration was developed. A plot of this calibration for one of our plastic sinctillators can be seen in Figure C.7. As expected, this calibration is very linear.

What was causing the problems with our simulated energy spectrum then? Well, as it turned out, the error was not with the simulation but with the measurement. During the course of the Compton spectrometer experiment, it was determined that the digitizer configuration file we had been using was incorrectly written. Due to a misunderstanding in the order of operations, the dynamic range had been severely compromised, leading to signal saturation at fairly low energies. Because the signals

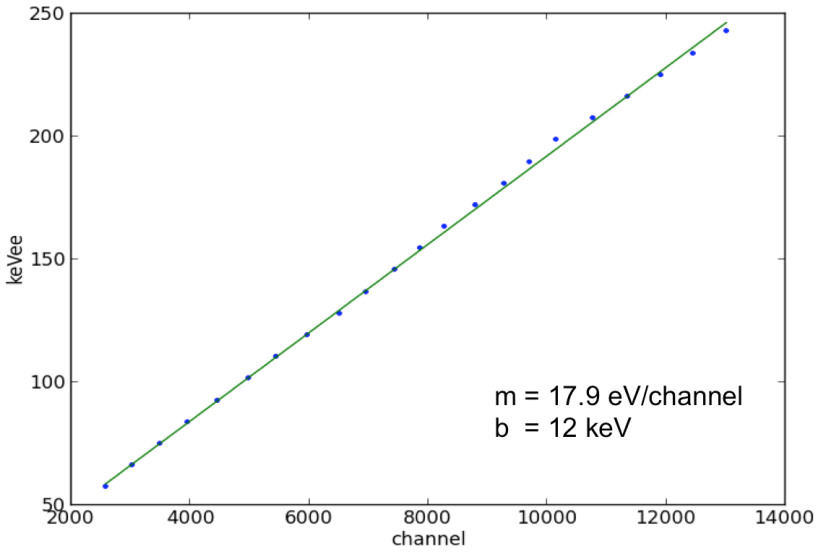


Figure C.7: The energy calibration measured for one plastic detector at a particular bias voltage.

are so peaked, however, the saturation effect was small enough to only cause a slight nonlinearity at higher energies. This only slightly distorted spectral features, allowing the incorrect data to pass visual inspection.

It was also determined that scattering from the environment was not adequately represented in our simulations. Previous figures show that simulations underestimated the spectrum in low energy regions. In a measured spectrum, room scatter and scatter from materials near the source makes up a significant portion of the measured spectrum. To measure the contribution from room scatter, measurements were taken with and without a lead shield in front of the detector. The measurement with a lead shield would block direct interactions from the source, allowing only room scatter interactions. Figure C.8 shows the results of these measurements in a plastic scintillator. In these measurements, room scatter accounted for approximately one fifth of the detected signal.

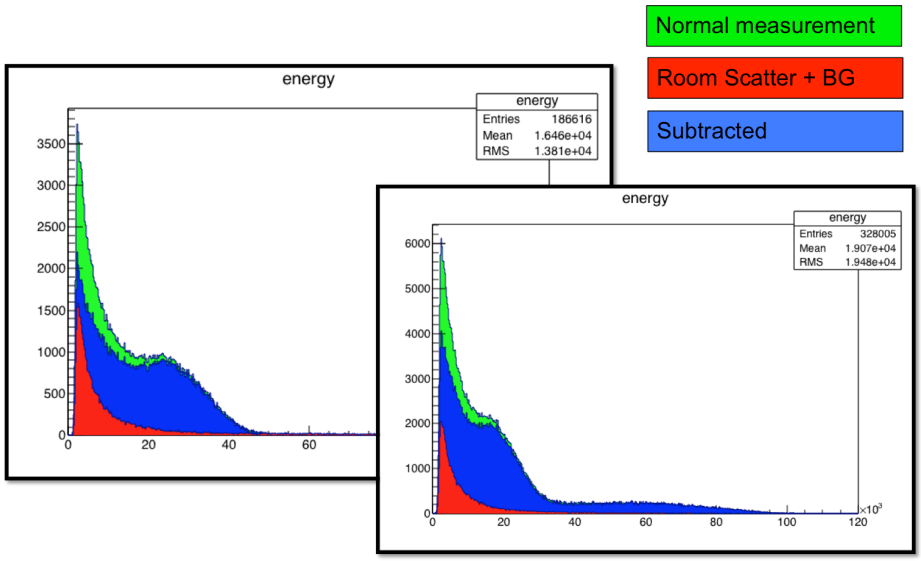


Figure C.8: Measured room scatter for both Cs-137 and Na-22 in a plastic scintillator. Measured room scatter is shown in red, while a measurement including room scatter is shown in green. Their difference, shown in blue, is the expected energy spectrum without room scatter. This subtracted spectrum is a much closer match to our simulated results.

Appendix D

PSD Data

Pulse shape discrimination (PSD) in the DDLI PoC was accomplished with a modified version of the NGMPSDMaker class from the NGM data readout and analysis library. NGMPSDMaker analyzed measurements taken in the trailer with Cf-252 pellets to fit the neutron and gamma bands in the resulting PSD vs. energy plots with cubic splines. These cubic splines represented an energy dependent mean and standard deviation for the gamma and neutron bands. Knowledge of these quantities was later used to perform PSD in the mobile measurements.

The splines can be used to calculate a figure of merit for PSD in the organic scintillators according to the formula

$$FOM = (\mu_g - \mu_n)/(\sigma_g + \sigma_n) \tag{D.1}$$

The results of this formula at different energies are shown in Tables D.1, D.2, and D.3. These values give an indication of PSD performance in the individual PoC detectors.

Table D.1: A figure of merit for PSD evaluated at different energies in plastic scintillators 1-10 of the DDLI PoC.

Detector	100 keVee	200 keVee	662 keVee
Plastic 1	0.98	1.76	3.32
Plastic 2	1.06	1.61	2.01
Plastic 3	1.05	1.80	3.30
Plastic 4	1.09	2.02	3.68
Plastic 5	1.20	2.23	3.64
Plastic 6	1.15	2.22	3.72
Plastic 7	1.06	2.01	3.73
Plastic 8	1.16	2.25	3.71
Plastic 9	1.07	1.72	3.43
Plastic 10	1.02	1.75	3.54

Table D.2: A figure of merit for PSD evaluated at different energies in plastic scintillators 11-20 of the DDLI PoC.

Detector	100 keVee	200 keVee	662 keVee
Plastic 11	0.97	1.81	3.47
Plastic 12	1.04	1.49	1.85
Plastic 13	0.71	1.48	3.33
Plastic 14	0.73	1.20	1.60
Plastic 15	0.71	1.28	1.72
Plastic 16	1.11	2.21	3.12
Plastic 17	0.92	1.84	3.52
Plastic 18	1.08	1.57	1.79
Plastic 19	0.89	1.71	3.50
Plastic 20	0.88	1.57	3.47

Table D.3: A figure of merit for PSD evaluated at different energies in the liquid scintillators of the DDLI PoC.

Detector	100 keVee	200 keVee	662 keVee
Liquid 1	0.69	1.15	1.53
Liquid 2	0.70	1.35	2.26
Liquid 3	0.74	1.61	3.26
Liquid 4	0.64	1.23	3.45
Liquid 5	0.55	1.37	3.73

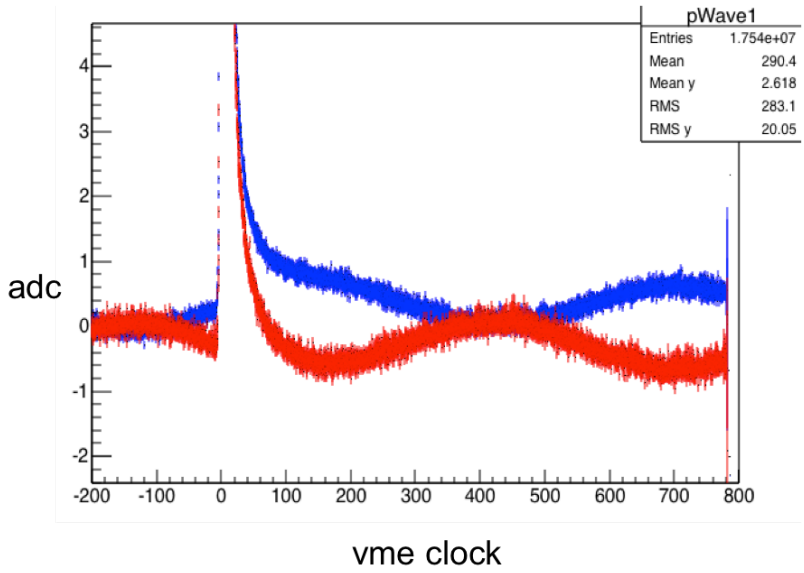


Figure D.1: Subsample noise in the liquid scintillators widens the gamma and neutron bands.

D.1 Discussion

Some detectors, such as Plastic 12 and Liquid 1, exhibit lower PSD performance than others. Several factors may have contributed to the low performance of these detectors. Figure D.1 shows average waveforms from either extreme of the gamma band of a liquid scintillator. The liquid scintillators were susceptible to sub-sample noise, as can be seen here. While the sinusoidal noise is lower in amplitude than even a single ADC unit, random fluctuations mean it still contributes to the waveform shape. Because this sinusoidal noise is unrelated to the waveform, it is not guaranteed to be in phase. Differences in the phase of this sinusoidal noise serve to widen the gamma and neutron bands in the liquid scintillators that are affected, reducing PSD performance.

In addition to this subsample noise, an unusual structure was also observed in some PSD plots. Figure D.2 shows one such PSD plot exhibiting three bands instead of two. The topmost band is of unknown origin. This band has an extremely high tail-to-total (PSD) ratio of close to unity, indicating fast pulses. The abnormal third

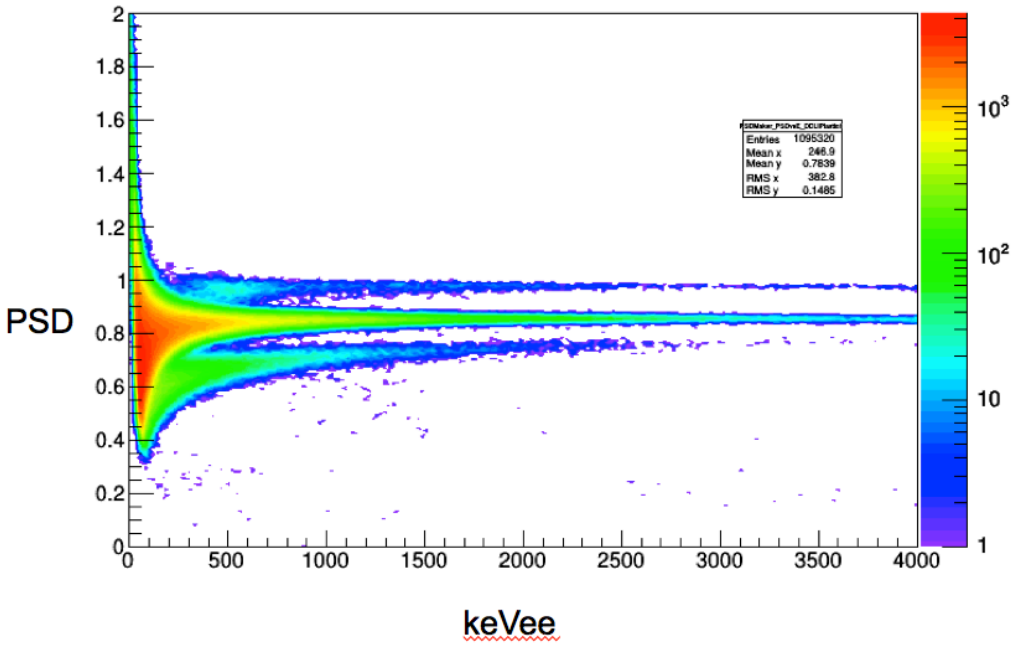


Figure D.2: Some of the organic scintillators exhibited a third band in their PSD plots. The cause of this third band is unknown.

band was seen in a subset of the plastic and liquid scintillators. The exact cause of the third band remains undetermined. Because NGMPSDMaker expects two bands only, the extra band likely degrades the fits and reduced PSD performance.

Appendix E

Measurement Plan Document

The Dual Detection-Localization-Identification (DDLI) system is a mobile imaging platform consisting of radiation detectors and a lead/polyethylene coded aperture mask in a fully enclosed trailer. In order to collect data for the validation of simulations, source measurements must be taken with both gamma ray and neutron sources. The DDLI prototype system validation measurements will take place at the parking lot next to the intersection of Bethel Valley Road and Spallation Drive. The measurements require that the imager be towed very slowly past stationary sources placed a significant distance away from the system (30 meters). The source will be stationed either in the parking lot or in the field on the opposite side of the road depending on the standoff distance required.

The measurements will take place after hours over the course of several days (approximately a week of good weather). At least two Rad-Worker qualified persons will be present during the actual measurements, one of which must be Rad-Worker II qualified for the RWP. The first worker will drive the NSITD Ford truck towing the DDLI prototype trailer along the road for multiple slow-speed passes by a single source. The second worker, outside the vehicle, will observe the source for the duration

of the measurement to maintain source control. The source is expected to be placed anywhere from 10-50 meters away from the road in a roped off Radiation Area. The source will be unshielded and elevated 1-2 meters off the ground level by a stand.

Note: At no time will anyone not associated with this project be exposed to radiation levels such that they might receive a whole body dose > 50 mrem.

The sources we plan to use in this experiment are the following Cs-137 source and Cf-252 source. *(See Table E.1.)*

Table E.1: Sources for use in DDLI PoC measurements. *(Note that the activity of the Cf-252 source was later measured to be lower than the calculated value listed here.)*

Nuclear Material Item#	Origin Date	Origin Assay [Ci]	Amount as of 06/12/2015 [Ci]
Cs-137-5658	06/01/2007	$1.01E-3$	$8.41E-4$
Cf-252-5557	03/13/2007	$9.96E-4$	$1.14E-4$

As of June 2015, these sources contain activity that is regulated by DOT. The Cs-137 source contains a DOT Limited Quantity amount of activity, while the Cf-252 source contains a Type A Quantity amount of activity. ORNL Onsite Transportation Safety Document specifies the following:

Radioactive sealed sources in less than Type B quantities transported or escorted by RCT technicians are excepted from TMO approval and the requirements of Section 3.8.

The sources will be transported to the measurement site by an RCT in a government vehicle. All sources used in this exercise will be leak tested prior to use in this field activity and have an ORNL Yellow Radioactive Material Tag.

When not in use, the sources will be stored in shielded configuration locked in the DDLI trailer by an experimenter. The source storage container will be labeled, Caution, Radioactive Material. The trailer will remain parked at the measurement site whenever a source is stored inside the trailer. As part of source control, the

presence of the sources in the trailer will be verified daily by an experimenter or RCT. Sources will be transported back to 3500 by the RCT in a government vehicle when the measurement is complete or the sources need to be exchanged.

Vita

John Sparger was born in Abingdon, VA in November 1987. He was raised by his loving mother, Dianna, and grandparents, Hilda and Harold Warren. After graduating from high school in the small town of Bristol, TN, John moved to Knoxville to begin his undergraduate studies. In 2010, John received his Bachelors of Science degree in Nuclear Engineering from the University of Tennessee. He then accepted a graduate research position at UTK with Dr. Jason Hayward. He was awarded a Masters of Science degree in December, 2012. Finally, in pursuit of his doctorate, John began work on research projects at Oak Ridge National Lab. John received his PhD in Nuclear Engineering in May, 2016. He is currently working with the Nuclear Security and Isotope Technology Division at ORNL.